GPU SMITH
COMPUTE ENGINEERING

GPU SMITH / RESEARCH & PRACTICAL GUIDANCE

GPU Node Sanitization: Handoff, RMA and Retirement Proof

> Plan and build private AI compute with GPU Smith. We connect workload requirements with GPU hardware, networking, deployment and operating decisions so your infrastructure fits the work it must perform.

Published by GPU Smith · 2026-11-18 · gpu node sanitization checklist · gpu data sanitization · gpu server rma data security · gpu server decommissioning · tenant gpu handoff · nist sp 800-88 rev 2 · nvme secure erase · certificate of destruction · bare metal gpu · statement of volatility

Original article: https://gpusmith.com/articles/en/gpu-node-sanitization-handoff-rma-retirement



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes in the Guidance: What Counts as Evidence
4. Data Path Inventory: What a GPU Node Can Retain
5. Supported Methods by Component and Their Limits
6. Four Dispositions and the Release Decision
7. Implementation Considerations and Process Changes
8. Data Analysis and Evidence
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

A **GPU node sanitization checklist** has to answer one question before any server, GPU board or drive leaves a tenant boundary: is there accepted evidence that tenant data cannot be recovered from each data-bearing component, or must the asset be held? This report treats that as a release decision per component, not a single "wipe" step. The distinguishing point is that eligibility for diagnosis or manufacturer repair is a different permission from authority to transfer data-bearing parts.

Two current documents frame the evidence. **NVIDIA Requirements for AI Clouds v2.4** (revision dated September 1, 2026) says sanitization must happen between tenants or on hardware replacement, and lists cryptographic erase of data drives, wiping persistent or volatile memory "including SRAM and GPU memory," and resetting the TPM and BIOS (docs.nvidia.com). It binds the NVIDIA Cloud Partner relationship, not every private owner. **NIST SP 800-88 Rev. 2** (September 2025) separates verification, which checks that a technique completed without errors, from validation, in which the organization reviews that data against sensitivity and accepts residual risk (csrc.nist.gov). NIST also states that it refrains from endorsing specific product-related measures (csrc.nist.gov), so no standard supplies a universal GPU erase command.

The component evidence is uneven. For NVMe media, the drive itself reports sanitize status and a purged bit (man.archlinux.org). For GPU memory, NVIDIA documents a reset as clearing "GPU HW and SW state" and warns that it is not guaranteed to work in all cases. In the NVIDIA pages reviewed here, a scrub of "state and memory" is stated only for the H100 confidential computing flow (developer.nvidia.com). An 8-GPU DGX H100 class node lists 640 GB of GPU memory, 2 TB of system memory and 34.56 TB of local flash (derived from 2 x 1.92 TB plus 8 x 3.84 TB) (docs.nvidia.com), so most local capacity sits on drives where documented, loggable methods exist.

The practical rule is a coverage ledger: components with accepted evidence divided by inventoried in-scope components. That fraction measures documentation completeness, not the probability of secure erasure, and any unknown critical component blocks a whole-node "verified" label. Failed or unreadable components stay unverified, never "erased." Vendor retention services such as drive-retention options let an owner keep failed

media on replacement (www.hpe.com), but GPU boards are covered only under model-specific terms. This report is operational engineering guidance, not a compliance certification, and it names no universal command.

Introduction and Background

Private GPU operators reuse expensive nodes across tenants, send failed boards to manufacturers, and eventually retire them. Each movement carries the same risk: data may persist in places a normal reinstall never touches. This report is written for infrastructure operators, tenant security owners, IT asset managers and procurement or support teams who must decide whether to release an identified node, GPU board or storage component for internal reassignment, tenant handoff, manufacturer return or retirement, or to hold it until supported sanitization and evidence exist.

Control boundaries matter because they change the required assurance. NIST SP 800-88 Rev. 2 treats media turned over to a maintenance provider as potentially still under organizational control only if a contractual agreement provides for confidentiality (nvlpubs.nist.gov). It treats media exchanged for warranty, and not returned, as out of organizational control (nvlpubs.nist.gov). A repair return and a warranty exchange can therefore demand different evidence for the same board.

The scope of this report is deliberately narrow. The NVIDIA requirements discussed below are written for the NVIDIA Cloud Partner relationship, and they are a requirement set for that relationship, not a legal compliance guarantee for any other site. No source reviewed documents a procedure that sanitizes an entire GPU server in one operation, and the report does not invent one.

This report is written from an independent advisory perspective: it neither makes GPUs nor operates a cloud, and nothing here certifies a result. Where vendor documentation could not be opened or is silent, the report says so and treats the component as unknown. Dates are anchored to November 2026, and most vendor pages carry no revision date, so each procedure must be rechecked against the exact SKU and firmware in the inventory.

Key Changes in the Guidance: What Counts as Evidence

From technique tables to a sanitization program

NIST published SP 800-88 Rev. 2 in September 2025, superseding Rev. 1 of December 17, 2014 (csrc.nist.gov). NIST's own FAQ describes the shift: Revision 1 focused on media-specific techniques in tables, while Revision 2 emphasizes an enterprise program, sanitization assurance, modern logical storage and a decision flow based on factors such as sensitivity and reuse inside or outside the organization. The FAQ confirms that Revision 2 does not recommend specific techniques for each type of media and points instead to evolving industry standards such as IEEE 2883 (csrc.nist.gov).

Two consequences follow for GPU nodes. First, the standard defines sanitization as rendering access to target data infeasible for a given level of effort (csrc.nist.gov), so the target is the data on each storage location, not the word "GPU." Second, NIST states that multi-pass overwriting is unnecessary for modern storage (csrc.nist.gov), which affects how long a drive step should take but says nothing about GPU memory.

Verification and validation as two separate acts

The central change for evidence is the split between two roles. Verification inspects whether the technique completed, and for clear and logical purge it can be done by checking tool completion status, errors, anomalies and the health of the media (nvlpubs.nist.gov). Validation is the decision that the target data was effectively sanitized, which can approve the result or reject it and require repetition or escalation (nvlpubs.nist.gov).

NIST lists conditions that can defeat a successful-looking operation: media that appears functional but has inaccessible regions, a method mismatched to the media, unqualified personnel or uncalibrated tools, and a scope that was too narrow, such as simple overwriting of overprovisioned media (nvlpubs.nist.gov). It also notes that elaborate sampling of media contents after clear or purge is not necessary unless organizational policy requires it (nvlpubs.nist.gov). The owner of the data, not the tool, signs the acceptance.

The NVIDIA requirement: between tenants and on replacement

The NVIDIA AI Cloud requirements version cited here is v2.4. Requirement SEC21 reads that sanitization "must be performed between tenants or on a hardware replacement" (docs.nvidia.com). Requirement OPO03 adds that deletion on offboarding "shall be defined and verifiable" (docs.nvidia.com).

The storage requirement STG02 is the most operationally specific. It permits an optional flag to skip drive sanitization in break-fix flows where tenancy does not change, but states that critical hardware component replacements may require sanitization without an override, including GPU and CPU node local storage (docs.nvidia.com). That distinction maps directly onto the dispositions discussed below. The document is a requirement, not an implementation procedure: it names what must be covered and does not give a GPU-specific method.

A UEFI media-sanitization menu is not a node proof

NVIDIA's DGX Spark UEFI guide shows a device and method menu under Security, where the operator selects the storage device, then a method type of Clear or Purge, then a specification. The same page documents Secure Boot key management separately. The menu operates on a selected storage device, so its existence does not establish coverage of GPU or host memory or whole-system disposal, and its steps must not be generalized to other platforms such as DGX B300.

Data Path Inventory: What a GPU Node Can Retain

The first deliverable of any release decision is an inventory built from the actual node, not from a catalog description. NIST observes that many items contain several types of media needing different methods, and that a statement of volatility (SoV, a manufacturer specification listing volatile and non-volatile memory locations, data nature, user accessibility and erase method) can inform decisions, with the caution that vendors state volatility details differently (nvlpubs.nist.gov). *Table 1* below lists data-bearing candidates using documented DGX-class configurations as the example; each row must be confirmed against the OEM statement of volatility and the real bill of materials.

Component	Possible data class	Volatility or persistence claim (source)	Documented method or gap	Prerequisite	Evidence to accept, and owner	Destination
GPU HBM (8 per node)	Tenant tensors, weights, activations	NVIDIA describes reset as clearing "GPU HW and SW state" (docs.nvidia.com)	Scrub of "state and memory" stated for the H100 CC flow; no equivalent found for non-CC reset	GPU model, CC mode, firmware version	Vendor statement for that SKU; otherwise UNKNOWN. Owner: security and data owner	Reassign only after acceptance
GPU persistent records (InfoROM, row remap)	ECC history, retired pages, configuration	"persistent store of configuration and state data" (docs.nvidia.com)	No documented erase method found	Decide whether to retain as RMA history	Recorded as device history, not as tenant content, per vendor description	Stays with the board
Host DRAM (2 TB, 32 DIMMs)	OS and process memory, key material	Volatile media lose contents when power is removed (nvlpubs.nist.gov)	Power-off; no minimum power-off duration is documented in the sources cited here, and keys may be unwrapped in memory	Confirm power state	Power-cycle record; owner: platform	Any
OS and boot NVMe (2 x 1.92 TB)	OS, logs, credentials, cached tenant data	Non-volatile; boot drives are two 1.92 TB M.2 devices in the DGX guides	Method and key custody differ by drive model	Drive model, firmware, key custody	Sanitize log status plus serial; owner: data owner	Reuse or hold
Local data and cache NVMe (8 x 3.84 TB)	Datasets, checkpoints, scratch	Self-encrypting, commonly RAID 0 (docs.nvidia.com)	Crypto erase replaces the key so earlier data is unintelligible (www.micron.cn); Dell states reassigned blocks are scrambled too (www.dell.com)	Key inventory, no off-node key copies	Per-drive evidence; owner: data owner	Reuse, retain or destroy
TPM 2.0	Wrapped SED vault keys, platform secrets	Non-volatile; vault data persists even if reimaged (docs.nvidia.com)	TPM clear removes owner authorization and stored keys (learn.microsoft.com); HPE reports clearing TPM nonvolatile data (support.hpe.com)	Know which keys depend on it	Clear record; owner: platform	Reuse
UEFI or BIOS NVRAM and boot keys	Passwords, boot order, Secure Boot keys	Non-volatile settings	Redfish action to clear the BIOS on DGX H100 (docs.nvidia.com)	Platform-specific	Action log; owner: platform	Reuse
BMC (configuration, users, logs)	Credentials, network config, IPMI, audit, video and POST logs	A Dell SoV lists the iDRAC SPI flash as non-volatile (dl.dell.com)	Factory reset and log erase (redfish.dmtf.org) (www.dell.com)	Preserve evidence before erase	Reset record; owner: security	Reuse

Component	Possible data class	Volatility or persistence claim (source)	Documented method or gap	Prerequisite	Evidence to accept, and owner	Destination
NICs and DPUs	Firmware configuration, DPU eMMC and NVMe	BlueField factory reset covers the BMC, UEFI, NIC and Arm (docs.nvidia.com)	DPU storage wipe is a separate step (docs.nvidia.com)	Confirm which devices are fitted	Per-device record	Reuse

The matrix shows why one command cannot be a whole-node proof. Only the drive rows have standardized, device-reported completion evidence; the GPU memory row is largely a documentation gap. Rows marked UNKNOWN are not failures of the process. They are the correct recorded state until vendor documentation or a supported method is obtained.

GPU and host memory: reset, scrub and persistence

NVIDIA's `nvidia-smi` documentation says a GPU reset can clear GPU hardware and software state that would otherwise need a reboot, but that reset is not guaranteed to work in all cases (docs.nvidia.com). A complete reset after an unhealthy result is a node power cycle (docs.nvidia.com). The pages describe state, not erasure of framebuffer contents. In the pages read for this report, NVIDIA does not state that HBM is zeroed on plain reset, driver unload or power cycle; that is a silence finding, not evidence of a problem.

In the pages reviewed, the initialization-time ECC scrub is described only as a latency effect, not as a sanitization step, so it should not be cited as one. Confidential computing documentation is more explicit: the trust model says GPU reset and teardown mechanisms clear protected state (docs.nvidia.com), yet also that stopping a workload is not a substitute for that lifecycle (docs.nvidia.com).

Persistent device history versus tenant content

Board-lifecycle items such as ECC mode, aggregate ECC counts and retired pages last across boot cycles (docs.nvidia.com). These are health records. They are also the history an RMA triage depends on, which is why the release record should classify them as retained diagnostic evidence rather than data to destroy. No NVIDIA statement reviewed says the InfoROM holds tenant content, and none describes an erase method for it.

Off-node copies and diagnostic bundles

Diagnostic output leaves the node. The legacy NVIDIA RMA submission document, hosted by a third party, dated March 2014 and used here as historical context only, describes a bug-report script whose output includes information such as `lspci` and system message log files and states that logs should accompany any submission (public.dhe.ibm.com). The current NVIDIA RMA page that earlier guides pointed to returns "Page Not Found," so current SKU or OEM instructions and approved redaction rules must be obtained. DCGM diagnostics write debugging logs and the memory test writes patterns to GPU memory (docs.nvidia.com), so the order of diagnostics and sanitization matters.

Supported Methods by Component and Their Limits

Drives: Sanitize, Format, Opal Revert and crypto erase

For NVMe media the standards are mature; IEEE 2883 covers methods of sanitizing logical and physical storage (standards.ieee.org), and the UK NCSC names ATA and NVMe Sanitize commands as manufacturer-provided overwrite aids (www.ncsc.gov.uk). NVM Express describes Sanitize as the other command to erase user data besides Format, and notes that it restarts automatically after a device reset (nvmexpress.org). The sanitize log reports whether the subsystem has never been sanitized, whether the latest operation completed successfully and whether it purged data as defined by IEEE 2883 (man.archlinux.org). A "Global Data Erased" bit indicates whether the drive was written since the last successful sanitize (man.archlinux.org), which is useful post-sanitize evidence. Time estimates may not be reported; the value FFFFFFFFh means none (www.mankier.com).

Caveats are documented. A SNIA primer states that overprovisioned areas are not accessible through the drive interface and that the interface does not provide observability of everything sanitized (www.snia.org). A second SNIA deck notes that crypto or block erase can invalidate media ECC, which complicates read-back verification, and lists namespace-level sanitization as ongoing work (www.snia.org). The same deck says documentation of sanitization is, or should be, required (www.snia.org). TCG Opal behaves differently: a successful Revert eradicates media keys, but an interrupted Revert is aborted and the result of user data removal is undefined (trustedcomputinggroup.org), and an SP still in the Manufactured-Inactive state is not affected (trustedcomputinggroup.org). Drives without encryption, such as Pyrite class, do not support cryptographic scramble (www.micron.cn).

Keys: scope, copies and dependencies

Crypto erase moves the problem from data to keys. NIST's FAQ lists preconditions: FIPS-validated modules for federal agencies, no prior plaintext on the media, adequate strength, and permanent zeroization of the target keys (csrc.nist.gov). NIST also warns that if target keys or keys below them exist outside the media, for example from a key management server or escrow, the assumption weakens (nvlpubs.nist.gov). A key inventory should therefore list:

- **Drive media keys:** held in the drive, erased by the vendor method; confirm the drive actually uses them (non-encrypting drives do not).
- **Authentication keys in the TPM:** DGX documentation says that after clearing the TPM the vault and SED authentication keys must be re-initialized (docs.nvidia.com).
- **External key management:** Dell warns that with SEKM enabled the feature must be disabled before System Erase or iDRAC-secured devices can be permanently locked (www.dell.com).
- **Recovery credentials:** a drive PSID revert makes data inaccessible and returns the drive to factory settings (solidigm.com).
- **Tenant-held secrets and attestation:** the tenant should re-attest after resets before authorizing secret release (docs.nvidia.com).

OEM whole-system erase: scope differs by vendor

OEM features cover more than drives but differ in scope. Dell's System Erase has separate switches for BIOS, iDRAC, Lifecycle Controller data and drives, uses Instant Secure Erase for capable devices and overwrite for others, does not erase iDRAC rollback firmware, and records actions in the Lifecycle Log (www.dell.com). HPE's one-button erase skips drives that do not support purge, states that supported devices that fail and unsupported devices are not erased securely, and on some Gen10 firmware reports only part of the device set (support.hpe.com).

Lenovo's BMC factory reset does not affect UEFI settings (pubs.lenovo.com), and its drive erase does not erase drives if a UEFI admin password is set (pubs.lenovo.com). Supermicro's published BIOS secure-erase guide is dated July 12, 2021 and covers X11, X12, B11 and B12 boards (www.supermicro.com), and its IPMI utilities are not supported on X14 or H14 platforms and later (www.supermicro.com). Any secure-deletion procedure in NVIDIA's DGX guides applies to the system drives; the sources cited in this report do not show it covering GPU memory.

What each action can and cannot prove

Table 2 below separates the common actions by what they evidence. It is a reading aid, not a procedure.

Action	Can evidence	Cannot evidence by itself
GPU reset or driver unload	State cleared per vendor docs; completion of the reset	Framebuffer erasure on non-CC systems (not stated); non-GPU components
Node power cycle	Loss of volatile state under normal behavior	Persistent media, TPM, BMC or firmware state
Diagnostic pass (DCGM or field diagnostics)	Hardware health, RMA eligibility inputs; logs written	Sanitization; it writes patterns to memory, which is not an erase guarantee
Drive sanitize or crypto erase	Per-drive status, purged bit, serial and tool version (man.archlinux.org)	Other drives, off-node key copies, overprovisioned areas not exposed by the interface
OEM system erase	Listed components cleared per vendor scope (support.hpe.com)	Components the vendor excludes or skips, GPU memory, attached accelerators
Whole-node release proof	Accepted evidence for every in-scope component, plus validation sign-off	Nothing, unless every inventoried component has accepted evidence

The key takeaway is that completion of one row never substitutes for another. A passing diagnostic is a health statement, a reset is a state statement, and only per-component evidence reviewed by the data owner supports a transfer.

Four Dispositions and the Release Decision

Same-owner maintenance

When the same tenant or owner keeps the data, such as a power supply swap, STG02 recognizes that tenancy does not change and allows a skip flag for drive sanitization. The exception is critical component replacements, so the decision belongs in the record. Preserve diagnostic evidence under access control inside the owner's boundary.

New-tenant reassignment

Reassignment inside the operator crosses a tenant boundary even though the hardware stays in the building. This is the SEC21 case. The handoff requires sanitization evidence for drives, memory scope, TPM and BIOS, with persistent configuration reset to baseline. Reused nodes then need requalification, covered in the next section.

External repair and RMA

Eligibility for RMA is a support decision. NVIDIA Enterprise Support says an RMA begins only after an engineer confirms entitlement through troubleshooting (docs.nvidia.com). Entitlement to an RMA does not grant permission to transfer data-bearing parts. Dell's media sanitization statement tells customers with sanitization obligations to ensure data has been removed, encrypted or sanitized before returning a device, and states that Dell has no contractual obligation to sanitize returned devices under warranty or support contracts (www.delltechnologies.com). HPE states that for leased, loaned or rented equipment the customer is solely responsible for removing sensitive data before return (www.hpe.com), and that replaced products become HPE property on deinstallation under its exchange service (www.hpe.com). Lenovo's agreement lists safeguarding or removing data before service among customer responsibilities (download.lenovo.com).

Retention options change the destination. Dell lets customers retain failed drives, and a component variant extends this to other parts (www.delltechnologies.com). HPE extends retention to failed GPU cards for supported models, with limits by GPU model and configuration (www.hpe.com). Lenovo limits its keep-your-drive service to non-volatile memory parts and requires serial numbers of each retained part (download.lenovo.com). The DGX B200 service manual says a purchased media retention policy may allow keeping the failed drive (docs.nvidia.com). Supermicro's RMA policy requires whole systems to ship in original pre-opened condition (www.supermicro.com), and the warranty page tells customers to ensure critical data is backed up (www.supermicro.com); the pages read for this report give no media-retention or wipe terms, so that handling is undocumented and must be asked.

Retirement

Retirement is the case where hold, retain or destroy decisions concentrate. The UK National Cyber Security Centre says media should be sanitised before the device leaves organisational control and that simply deleting is not enough (www.ncsc.gov.uk). NIST's FAQ adds that pulverize and shred techniques are ineffective for medium and high security categories because of data density (csrc.nist.gov), and that degaussing is a physical purge technique for legacy magnetic media, not a true destroy method (csrc.nist.gov). The NSA/CSS policy manual lists disintegration with a solid-state sanitization device approved for the device (media.defense.gov).

None of these sources describe destruction specifically of a GPU board, so the destruction method for HBM-bearing boards is a contract and policy decision to be made with the data owner, not a documented standard answer.

Failed or inaccessible hardware

A failed component must never be recorded as erased. NIST states that overwriting cannot be used on media damaged to the point of being inoperable (nvlpubs.nist.gov), and notes that media may not be in a usable state after a purge so inspection can be limited. *Table 3* below lists dispositions for common failed states.

State	Do not record	Candidate dispositions (policy or contract decision)	Evidence retained	Approver
Drive does not enumerate	"Sanitized" or "erased"	Hold; retain under a retention service; supported destruction	Serial, failure symptom, custody log	Data owner
Sanitize reports failure or is interrupted	"Purged"	Repeat with another technique, escalate, or destroy per NIST validation rejection	Status log, tool and version	Data owner
Device or method unsupported	"Verified"	Hold; apply a method supported for that device; destroy	Support statement, SKU, firmware	Security and asset owner
GPU board unresponsive	"Memory cleared"	Hold; vendor handling under a written agreement; retained-board policy	Serial, UUID, diagnostic bundle review	Data owner and procurement
BMC unreachable or credentials lost	"Logs and keys cleared"	Hold; manufacturer-supported recovery; retain under custody	Management-plane inventory	Security owner
Key custody unknown	"Crypto erased"	Treat as unverified until the key scope is established	Key inventory	Data owner

The table encodes one rule: when a method cannot run, the correct record is the state of the asset (held, retained, destroyed under evidence, or sent under an agreed vendor arrangement), never a fabricated proof. NIST makes the same point by allowing a validation decision to reject a result and require escalation.

Implementation Considerations and Process Changes

Handoff checklist

The checklist below is engineering guidance for a release decision. It contains no commands, because methods are version and SKU specific and must be approved by the asset and data owner.

1. **Name the disposition:** same-owner maintenance, new-tenant reassignment, external repair or retirement.
2. **Inventory the node:** record serials and UUIDs for every GPU, drive, NIC, DPU, TPM and management controller from the real configuration.
3. **Obtain volatility statements:** collect the OEM statement of volatility for each platform and mark unknown components as unknown.

4. **Freeze the evidence set:** preserve required diagnostic output under access control before any erase.
5. **Review diagnostic bundles:** inspect contents for host and tenant data, apply approved redaction, and decide what may be uploaded externally.
6. **Map keys:** list media keys, TPM-held keys, external key manager entries and recovery credentials.
7. **Select a supported method per component:** document who approved it and what happens on failure.
8. **Run per component:** capture tool, firmware version, serial, timestamps, anomalies and exit status.
9. **Verify:** check completion status and errors against the vendor-defined evidence source.
10. **Validate:** the data owner accepts or rejects against sensitivity, and records residual risk.
11. **Handle failures:** hold, retain or destroy; never mark failed or unreadable parts as erased.
12. **Record custody:** destination, carrier, receiver and date.
13. **Restore baseline and requalify:** for reused hardware, without restoring prior tenant secrets.

The asset-release record

NIST's sample certificate of sanitization records manufacturer, model, serial number, media type, source, method (clear, purge, destroy), technique, tool used including version, verification method and the people who verified and validated, with a destination checkbox including internal reuse, external reuse and manufacturer (nvlpubs.nist.gov). ADISA's standard sets a minimum certificate or log content and includes the serial number of the device (adisa.global). The R2v3 interpretation says simple operator-created spreadsheet records are not acceptable for software sanitization and that serial numbers should preferably be read by the software (sustainableelectronics.org). For destruction, i-SIGMA's manual calls for recording serial numbers of drives being destroyed (isigmaonline.org) and for a receipt or certificate reflecting unique identifiers (isigmaonline.org). A GPU release record should add GPU UUIDs, firmware versions, anomalies, approvals, destination and the coverage fraction described below.

Restoring baseline without restoring secrets

Reset variants differ in what they retain. A Dell iDRAC reset has one form that keeps user and network settings and another that removes them (www.dell.com), and Lenovo's BMC reset retains whatever the operator checks (pubs.lenovo.com). A baseline restore must therefore reapply approved settings from a controlled source, not from a prior tenant's backup. OpenStack Ironic documents hardware-specific steps such as resetting the iDRAC to a known good state, and says its default agent ships with only a minimal set of cleaning behavior (docs.openstack.org). For confidential computing nodes the next tenant should attest after resets before secrets are released.

Turnaround is staged, not fixed

Turnaround is the sum of measured stages: queue, supported operation, validation and custody. No universal wipe duration exists. HPE says its secure erase can take up to a day (servermanagementportal.ext.hpe.com), OpenStack Ironic notes that erasing can be lengthy unless fast hardware-assisted erasure is supported (docs.openstack.org), and Kioxia states that a crypto erase renders data unreadable in less than a second on its drives (americas.kioxia.com). Those are vendor statements for specific products, so use measured times from the actual fleet and treat resale recovery rates as unknown.

Contract questions for spares, RMA and end of lease

- **Retention:** does the support contract allow keeping failed drives, components and GPU boards, and for which models?
- **Sanitization responsibility:** who sanitizes returned or exchanged parts, and does the vendor state a contractual obligation?
- **Diagnostic data:** which logs are required, how are they redacted, and how long does the vendor retain them?
- **Spares pool:** are spares sanitized before they enter and after they leave the pool?
- **End of lease:** who holds the key custody record and the certificate of destruction?
- **Documentation:** will the manufacturer supply the current statement of volatility and method documentation for the exact SKU?

Data Analysis and Evidence

Inventory scale

A DGX H100 class guide lists 8 GPUs with 640 GB total GPU memory, 2 TB of DIMMs, two 1.92 TB boot drives and eight 3.84 TB SED data drives (docs.nvidia.com). Local flash is 2 x 1.92 TB plus 8 x 3.84 TB, or 34.56 TB (derived), which is 54 times the 640 GB of HBM and 17.3 times the 2 TB of DRAM. A DGX B200 lists 1,440 GB of GPU memory. The DGX B300 guide lists 8 x 288 GB, or 2.3 TB (docs.nvidia.com), while NVIDIA's web page lists 2.1 TB (www.nvidia.com). Neither source explains the difference, so the inventory should record the figure measured from the installed node.

Guidance cadence

NVIDIA's requirements document moved through four versions in about six months: 43 days from v2.1 to v2.2, 76 days to v2.3 and 68 days to v2.4, using the dates in its revision table (docs.nvidia.com). NIST's Rev. 2 final page dates from September 2025 and its FAQ from July 16, 2026. A release record should therefore cite the version in force on the day of release.

Coverage arithmetic (Hypothetical Example)

Coverage equals components with accepted evidence divided by inventoried in-scope components. Consider a hypothetical B200-class node with 8 GPU boards, 10 NVMe drives, 2 DPUs, a TPM, UEFI NVRAM, a BMC and host memory, totaling 24 in-scope components. If 22 have accepted evidence and the 2 DPUs have none, coverage is 22 of 24, or 91.7%. This number is a documentation completeness fraction, not a probability of secure erasure. Because two critical components are unknown, the node cannot receive a whole-node "verified" label, and the release is held or limited to the accepted components. Reaching 100% requires evidence, not rounding.

Implications and Future Directions

Several signals point toward more explicit multi-tenant evidence. IEEE lists a recommended practice for sanitizing before reuse, resale or disposal, and a draft covering multi-tenancy and self-encrypting devices (standards.ieee.org). Large providers describe their own approaches as vendor claims: AWS states it decommissions media using techniques in NIST 800-88 (aws.amazon.com), Google describes a move to cryptographic erasure after an earlier practice of overwriting drives with zeros (cloud.google.com) (services.google.com), Microsoft describes NIST SP 800-88 clear guidelines in its datacenters (learn.microsoft.com), AWS documents key destruction for NVMe instance store data (docs.aws.amazon.com), and in bare metal mode its Nitro design prevents the main CPUs from updating firmware (docs.aws.amazon.com), and Lumen documents that deleting a bare metal server cryptographically erases its storage (docs.lumen.com). OpenStack Ironic documents that cleaning steps depend on the hardware and that misconfigurations can lead to data exposure (docs.openstack.org).

For private operators, two gaps remain. Vendors need to publish component-level statements for GPU memory and accelerator firmware, and operators need to decide in advance what happens to failed boards. The validation decision belongs to the data owner, and no external party certifies the outcome.

Frequently Asked Questions (FAQs)

How should a GPU server be sanitized between tenants?

Treat it as per-component release: inventory, preserve evidence, run supported methods for drives, TPM, BIOS, BMC and memory scope, verify, validate, and record. NVIDIA's cloud requirement lists cryptographic erase of data drives, wiping persistent or volatile memory and resetting TPM and BIOS for its partner relationship as described above.

Is a GPU memory wipe between customers documented?

For confidential computing, NVIDIA documents that reset and teardown clear protected state. For other configurations, the pages reviewed do not state that framebuffer contents are zeroed, so the status is unknown until the vendor states it.

What about GPU server RMA data sanitization?

RMA eligibility is not transfer permission. Some vendors state that they are not contractually obliged to sanitize returned devices (www.delltechnologies.com), so the owner must sanitize, retain or agree a handling arrangement first.

How should a GPU server be decommissioned and its data destroyed?

Sanitize reachable media, then apply a destruction method appropriate to the media and sensitivity, with verification of the result; NCSC notes that particle size should be verified after destruction (www.ncsc.gov.uk), and CISA advises using a professional facility that offers a certificate of destruction (www.cisa.gov).

Does NIST 800-88 cover GPU servers?

NIST SP 800-88 Rev. 2 covers media sanitization programs and does not recommend per-media techniques, nor endorse products. It informs the process, not a GPU command.

Is there a bare metal GPU tenant handoff checklist?

The handoff checklist above is the engineering version. Ironic's automated cleaning is an open-source example that defaults to removing disk data and fails cleaning if a device cannot be erased (docs.openstack.org).

How does NVMe secure erase on a GPU node work?

Use the drive's supported sanitize or crypto erase method for the specific model and read the sanitize log for status. Interrupted Opal reverts leave user data removal undefined (trustedcomputinggroup.org).

What goes on a certificate of destruction?

At minimum, unique identifiers such as serial numbers, method, tool, date and responsible people. NIST's certificate form lists these fields, and i-SIGMA expects unique identifiers on the receipt (isigmaonline.org).

Conclusion

The release decision for a GPU node is a ledger of per-component evidence, not a single wipe. Diagnostic and RMA eligibility, drive sanitization, key custody, persistent configuration and off-node copies each need their own record, and the data owner validates the result. Where a method is unsupported, a component has failed or documentation is missing, the honest state is hold, retain, or supported destruction under agreement, never a fabricated proof. Coverage fractions describe documentation completeness, and a single unknown critical component keeps the whole-node label from being applied. Operators should obtain current OEM statements for each SKU, agree contract terms for retention and returns before failures occur, and treat every procedure in this report as requiring approval by the asset and data owner.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://docs.nvidia.com/dsx/ncp/nvidia-requirements-for-ai-clouds/v2.4
2. https://csrc.nist.gov/files/pubs/sp/800/88/r2/final/docs/sp800-88r2-faq.pdf
3. https://man.archlinux.org/man/nvme-log-sanitize.1.en.raw
4. https://developer.nvidia.com/blog/confidential-computing-on-h100-gpus-for-secure-and-trustworthy-ai
5. https://docs.nvidia.com/dgx/dgxh100-user-guide/introduction-to-dgxh100.html
6. https://www.hpe.com/services/hpe/pdf/a00108652enw
7. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r2.pdf
8. https://csrc.nist.gov/pubs/sp/800/88/r2/final
9. https://docs.nvidia.com/deploy/nvidia-smi/index.html

10. https://docs.nvidia.com/dgx/dgxb200-user-guide/introduction-to-dgxb200.html
11. https://www.micron.cn/content/dam/micron/global/public/products/technical-marketing-brief/micron-ssd-security-features-tech-brief.pdf
12. https://www.dell.com/en-us/infohub/articles/instant-scramble-erase
13. https://docs.nvidia.com/dgx/dgx-el9-user-guide/managing-seds.html
14. https://learn.microsoft.com/en-us/powershell/module/trustedplatformmodule/clear-tpm
15. https://support.hpe.com/hpesc/public/docDisplay?docId=a00105236en_us&page=GUID-8A7C60B7-A3A1-42C6-BEA5-0252F945B7C9.html
16. https://docs.nvidia.com/dgx/dgxh100-user-guide/redfish-api-supp.html
17. https://dl.dell.com/manuals/all-products/esuprt_ser_stor_net/esuprt_poweredge/poweredge-r740xd2_Reference-Guide2_en-us.pdf
18. https://redfish.dmtf.org/schemas/v1/LogService.v1_9_0.json
19. https://www.dell.com/support/kbdoc/en-us/000226396/poweredge-how-to-view-or-clear-the-system-event-log
20. https://docs.nvidia.com/networking/display/bluefieldbmcv2510ltsu1/factory-reset.pdf
21. https://docs.nvidia.com/confidential-computing/virtual-machines/threat-model
22. https://docs.nvidia.com/deploy/driver-persistence/data-persistence.html
23. https://public.dhe.ibm.com/systems/support/system_x_pdf/tesla_rma_process_da-07017-001_v03.pdf
24. https://docs.nvidia.com/datacenter/dcgm/latest/user-guide/dcgm-diagnostics.html
25. https://standards.ieee.org/ieee/2883/10277
26. https://www.ncsc.gov.uk/guidance/secure-sanitisation-storage-media
27. https://nvmexpress.org?p=4737
28. https://www.mankier.com/2/nvme_sanitize_log_page
29. https://www.snia.org/sites/default/files/SSSI/CMSI-Is-The-Data-Really-Gone-webcast-6-29-22.pdf
30. https://www.snia.org/sites/default/files/SSSI/CMSS24/CMSS24-Suhler-Storage-Sanitization-Recent-Evolution.pdf
31. https://trustedcomputinggroup.org/wp-content/uploads/TCG-Storage-Opal-SSC-v2.30_pub.pdf
32. https://www.dell.com/support/kbdoc/en-us/000460243/power-edge
33. https://solidigm.com/support-page/trouble-error-msgs/ka-00062.html
34. https://docs.nvidia.com/confidential-computing/bare-metal
35. https://pubs.lenovo.com/xcc2/NN1ia_c_resettingthexcc
36. https://pubs.lenovo.com/lxca/server_secureerase_drives
37. https://www.supermicro.com/manuals/other/X11_X12_B11_B12_Secure_Erase_Guide.pdf
38. https://www.supermicro.com/en/solutions/management-software/ipmi-utilities?mlg=0
39. https://docs.nvidia.com/enterprise-support-and-services-user-guide/nvidia-enterprise-support-details/index.html
40. https://www.delltechnologies.com/asset/de-de/solutions/business-solutions/legal-pricing/customer-facing-media-sanitization-statement.pdf
41. https://download.lenovo.com/km/media/attachment/Global%20Lenovo%20Infrastructure%20ISG%20Services%20Agreement%20-%20Version%207.4%20-%20September%202026%20-%20EN.pdf
42. https://www.delltechnologies.com/resources/en-us/asset/data-sheets/services/kyhd-kyc-for-enterprise-datasheet.pdf

43. https://docs.nvidia.com/dgx/dgxb200-service-manual/m2-nvme-replacement.html
44. https://www.supermicro.com/en/support/rma
45. https://www.supermicro.com/en/support/warranty
46. https://media.defense.gov/2026/Mar/10/2003897016/-1/-1/0/PM%209-12%2020260219.PDF
47. https://adisa.global/wp-content/uploads/2021/06/ADISA-Data-Capability-Requirements-Standard-8.0.pdf
48. https://sustainableelectronics.org/wp-content/uploads/R2v3-Formal-Interpretation-1.0-Data-Sanitization-Software.pdf
49. https://isigmaonline.org/wp-content/uploads/2026/07/i-SIGMA-Certification-Specifications-Reference-Manual_1026O-Edited.pdf
50. https://www.dell.com/support/kbdoc/en-us/000126703/poweredge-how-to-reset-the-integrated-dell-remote-access-controller-idrac
51. https://docs.openstack.org/ironic/latest/admin/cleaning.html
52. https://servermanagementportal.ext.hpe.com/docs/redfishservices/ilos/supplementdocuments/securesystemerase
53. https://americas.kioxia.com/content/dam/kioxia/en-us/business/ssd/asset/productbrief/KIOXIA_Enterprise_DataCenter_SSD_Security_Tech_Brief.pdf
54. https://docs.nvidia.com/dgx/dgxb300-user-guide/introduction-to-dgxb300.html
55. https://www.nvidia.com/en-eu/data-center/dgx-b300.md
56. https://aws.amazon.com/compliance/data-center/data-layer
57. https://cloud.google.com/blog/products/identity-security/the-future-of-media-sanitization-at-google
58. https://services.google.com/fh/files/misc/data_deletion_on_gcp.pdf
59. https://learn.microsoft.com/en-us/compliance/assurance/assurance-data-bearing-device-destruction
60. https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/data-protection.html
61. https://docs.aws.amazon.com/whitepapers/latest/security-design-of-aws-nitro-system/the-components-of-the-nitro-system.html
62. https://docs.lumen.com/edge-bare-metal/deleting-a-bare-metal-server
63. https://www.cisa.gov/resources-tools/training/protect-physical-security-your-digital-devices

THE PUBLISHER

About GPU Smith

GPU Smith provides independent engineering and research for private AI infrastructure. We help technical buyers and operators reason about GPU workload sizing, hardware selection, procurement, deployment and inference operations, from the compute system to the networking, power and cooling requirements around it.

Start with the workload

Useful infrastructure decisions begin with the models, data, throughput, latency and operating constraints a team actually has. GPU Smith helps connect those requirements with system design and deployment choices. The goal is a privately controlled AI environment whose capacity and operational demands are understood before a purchasing decision.

Hardware and supplier research

Our public reference library covers GPUs, complete systems and networking components. The server-vendor directory and manufacturing research help teams investigate suppliers, compare options and follow sources behind technical claims. We distinguish manufacturer specifications from measured benchmarks and advertised capabilities from tested configurations. Reference pages are research resources, not a live inventory listing or a binding equipment quote.

Deployment and operations

GPU Smith's engineering scope includes infrastructure integration, networking, power, cooling and the practical operation of inference workloads. Our published articles explain the tradeoffs behind deployment, procurement and ongoing operations so teams can ask better questions and document decisions.

Work with GPU Smith

Explore the hardware library, GPU references, networking references and vendor research. Contact GPU Smith with your workloads and deployment constraints to discuss a project and confirm current scope, pricing and availability.

Inclusion of a manufacturer or product in our research does not imply a partnership, certification or endorsement.

Website: https://gpusmith.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.