SYMBEEO

PUBLIC

DESIGN STUDY

Pharma research cluster: 32 GPUs, 2 PB of flash, InfiniBand fabric

Complete design on current platforms

CAPABILITY	INDUSTRY	SETTING
AI & Data	Life Sciences	Pharmaceutical research site
DOCUMENT	**VERSION**	**DATE**
AI-003	1.3	Oct 10, 2026

WRITTEN BY

Diksha Shah, Principal AI Infrastructure Architect

© 2026 Symbeeo. All rights reserved.

Document control

Document	AI-003	Version	1.3
Issued	Oct 10, 2026	Status	Issued by Symbeeo
Classification	Public	Supersedes	1.2, Sep 27, 2026
Capability	AI & Data	Industry	Life Sciences
Author	Diksha Shah	Role	Principal AI Infrastructure Architect
Credentials	NVIDIA Certified Professional: AI Operations, NVIDIA Certified Associate: AI Infrastructure and Operations, Google Cloud Professional Cloud Architect, AWS Certified Data Engineer		
Approval	Approved by Edwin Ozer, Founder, Oct 10, 2026. Issued by Symbeeo, Oct 10, 2026. Design study: every figure is vendor-published, a design target or a marked assumption.		
Sign-off	"An engagement under this design is signed by the practice lead before any build proceeds."		

Contents

1 · Scope 3
1c · Public implementations this design follows 4
2 · Requirements register 4
3 · Architecture overview 5
4 · Design by subsystem 6
5 · Failure behavior 7
6 · Limits 7
7 · Lessons learned 7
8 · The author 7

© 2026 Symbeeo. All rights reserved. · Public

CONVENTIONS

Source tags. [S] figure read from the vendor's own current page on Sep 18, 2026 · [S*] figure from the vendor's page summary where the page itself could not be opened · [A] design assumption, to be confirmed by the signing engineer · [T] design target taken from the study record or derived from it. No figure comes from any prior engagement, file or asset. A row with more than one tag is counted once: as a published figure if any tag is [S] or [S*], otherwise as an assumption if any tag is [A], otherwise as a design target.

Table 1. Revision history

Date	Version	Change
Sep 18, 2026	1.0	First issue.
Sep 20, 2026	1.1	Public implementations added; figures reconciled with published sources.
Sep 27, 2026	1.2	Figures and platform details rechecked against published sources.
Oct 10, 2026	1.3	Approval line now reads: approved by the Founder, issued by Symbeeo. Diksha Shah remains the author.

1 · Scope

Who it is for. The head of research computing at a pharmaceutical site whose six computational groups queue for GPU time on aging nodes, who needs one shared cluster that is fair between groups, fast enough for molecular and protein models, and auditable for the data it holds.

Typical buyer [A]. A pharmaceutical research site with six computational groups: 6 groups of 8–25 researchers, 4 principal workloads (protein structure prediction, molecular dynamics, image analysis, small-molecule generative models), 1.2 PB of active data today growing 30 % a year, one on-site data room with direct liquid cooling available.

What is in. Four 8-GPU nodes; a 400 Gb/s InfiniBand compute fabric; 2 PB of NetApp flash on a 200 GbE storage network; a job scheduler with fair-share across the six groups; the failure behavior. **What is out.** Model serving to the wider company; the wet-lab instrument network; identity and access design (SEC design studies); backup and archive tiers.

Assumptions. HPE is the server vendor, NVIDIA the accelerator and fabric vendor, NetApp the storage vendor [A]. Direct liquid cooling is available at the rack [A]. The scheduler is Slurm with fair-share accounting, one account per group [A].

Related documents, read Sep 18, 2026. HPE ProLiant Compute XD685 product page (https://www.hpe.com/us/en/products/compute/proliant/xd685.html) · HPE ProLiant Compute XD685 Direct Liquid Cooling Server, HPE Store (https://buy.hpe.com/us/en/Compute/Rack-Servers/Proliant-Compute-XD600-Servers/Proliant-Compute-XD600-Server/HPE-ProLiant-Compute-XD685-Direct-Liquid-Cooling-Server/p/S4R96A) · NVIDIA QM97XX 1U NDR 400 Gb/s InfiniBand switch systems user manual, introduction (https://networking-docs.nvidia.com/qm97x0hw/introduction) · NVIDIA Quantum-X800 platform page (https://www.nvidia.com/en-us/networking/products/infiniband/quantum-x800/) · NVIDIA Q32xx and Q34xx XDR switch systems user manual (https://networking-docs.nvidia.com/xdrswitcheshw/introduction) · NetApp AFF A1K key specifications (https://docs.netapp.com/us-en/ontap-systems/a1k/overview.html) · NetApp AFF A-Series page (https://www.netapp.com/aff-series/) · NVIDIA DGX SuperPOD with DGX B300 systems, Quantum-X800 InfiniBand and AC power reference architecture, read Sep 20, 2026 (https://docs.nvidia.com/dgx-superpod/reference-architecture/scalable-infrastructure-b300-xdr/latest/dgx-superpod-architecture.html) · NetApp NVA-1173, NetApp AIPod with NVIDIA DGX systems, read Sep 20, 2026 (https://docs.netapp.com/us-en/netapp-solutions-ai/infra/ai-aipod-nv-intro.html).

© 2026 Symbeeo. All rights reserved. · Public

Read Sep 27, 2026. NVIDIA QM97XX user manual, introduction (last updated Feb 16, 2026): 32 OSFP interfaces, 64 NDR 400 Gb/s ports, 51.2 Tb/s [S] · NVIDIA HGX AI Factory enterprise reference architecture (last updated May 18, 2026): HGX B300 systems with "NVIDIA ConnectX-8 HCA (on baseboard)" [S] (`https://docs.nvidia.com/enterprise-reference-architectures/hgx-ai-factory/latest/index.html`) · NVIDIA ConnectX-8 SuperNIC user manual, specifications (last updated Sep 7, 2026): InfiniBand XDR 800 Gb/s, NDR 400 Gb/s and NDR200 per port [S] (`https://networking-docs.nvidia.com/connectx8hw/specifications`) · NVIDIA InfiniBand adapters page: ConnectX-7 400 Gb/s, ConnectX-8 SuperNIC up to 800 Gb/s [S] (`https://www.nvidia.com/en-us/networking/infiniband-adapters/`) · HPE InfiniBand XDR Adapters QuickSpecs: HPE's XDR adapters are based on the ConnectX-8 SuperNIC [S*], HPE's pages could not be opened, published summary used (`https://www.hpe.com/psnow/doc/a50009229enw`) · NVIDIA Q32xx and Q34xx user manual (last updated Aug 21, 2026): Q3200-RA 72 XDR ports over 36 OSFP cages, 57.6 Tb/s [S].

1c · Public implementations this design follows

Deployment rows are real public deployments, each dated. Where no public deployment matches, the row is Vendor documentation: the vendor's document that this design, Symbeeo's own, is built on.

Table 2. Public implementations this design follows

Implementation	Who published it · date	What it settles for this design	URL
Vendor documentation. NVIDIA DGX SuperPOD with DGX B300 systems on Quantum-X800 InfiniBand, AC power	NVIDIA · last updated Sep 2, 2026	Blackwell Ultra 8-GPU nodes on an InfiniBand compute fabric; four B300 systems per rack at about 56 kW, the density this design stays under	`https://docs.nvidia.com/dgx-superpod/reference-architecture/scalable-infrastructure-b300-xdr/latest/dgx-superpod-architecture.html`
Vendor documentation. NetApp NVA-1173, NetApp AIPod with NVIDIA DGX systems	NetApp · Sep 15, 2025	NetApp AFF as the shared storage behind 8-GPU NVIDIA nodes, the pairing in R9	`https://docs.netapp.com/us-en/netapp-solutions-ai/infra/ai-aipod-nv-intro.html`

2 · Requirements register

Table 3. Requirements register

#	Requirement	Value	Source
R1	GPU count and layout	32 GPUs across 4 nodes, 8 per node	[T]
R2	GPU generation	NVIDIA Blackwell Ultra, HGX B300, 8 per XD685	[S] HPE Store XD685
R3	Node form factor	5U direct liquid cooled; 6U air	[S] HPE Store XD685
R4	Node CPU and memory	2 × 5th Gen AMD EPYC; 24 × DDR5-6400 RDIMM	[S] HPE Store XD685
R5	Compute fabric	400 Gb/s InfiniBand per port	[T]
R6	Fabric switch	64 × 400 Gb/s NDR on 32 OSFP cages in 1U, 51.2 Tb/s, managed	[S] NVIDIA QM9700

© 2026 Symbeeo. All rights reserved. · Public

#	Requirement	Value	Source
R7	Fabric ports per node	8 × 400 Gb/s NDR, one per GPU, on the node's eight ConnectX-8 adapters	[A]; [S] ConnectX-8 supports NDR 400 Gb/s, read Sep 27, 2026
R8	Flash capacity	2 PB usable	[T]
R9	Storage platform	AFF A1K: 14.688 PB raw per HA pair, 2 TB memory per pair, 2U per controller	[S] NetApp A1K
R10	Storage network	200 GbE, 24 ports per HA pair available	[S] NetApp A1K
R11	Storage scale-out ceiling	12 HA pairs, 176.3 PB raw (NAS)	[S] NetApp A1K
R12	Scheduling	fair-share across 6 groups, one account per group	[T]
R13	Job availability	99 % of scheduled node-hours delivered per month	[A]
R14	Growth	30 % data growth a year; +4 nodes in year 2 without fabric change	[A]

Fourteen rows; eight carry a vendor figure, four are design targets, two are design assumptions.

3 · Architecture overview

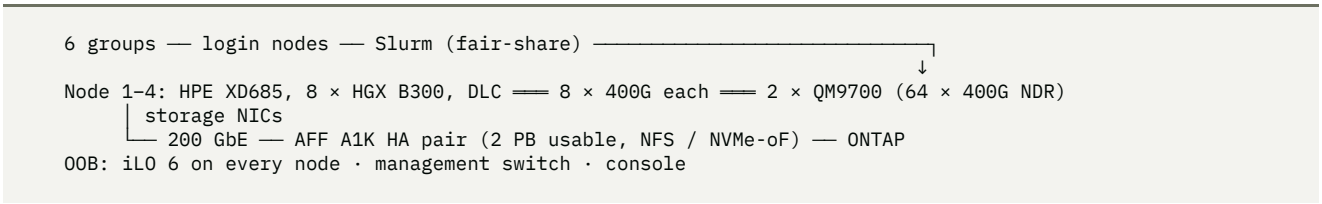


Figure 1. Research cluster, fabric and storage

Four identical 8-GPU nodes on a two-switch InfiniBand fabric with headroom for four more; one NetApp HA pair holds all active data; Slurm fair-share makes the six groups' allocations visible and enforceable.

Table 4. Architecture overview

Role	Product (exact model)	Qty	Capacity used	Why this unit
GPU node	HPE ProLiant Compute XD685, direct liquid cooling, 8 × NVIDIA HGX B300	4	32 of 32 GPUs [T]	Current HPE 8-GPU Blackwell Ultra node; DLC on all GPU options [S]
Fabric switch	NVIDIA Quantum-2 QM9700, 64 × 400 Gb/s NDR, 1U	2	16 of 64 ports each on day one; 32 of 64 after the R14 growth [T] derived (§4.2)	400 Gb/s per port (R5); onboard subnet manager [S]
Fabric adapter	NVIDIA ConnectX-8, up to 800 Gb/s, run at NDR 400 Gb/s, 8 per node	32	1 per GPU [A]	NVIDIA's HGX B300 reference places ConnectX-8 on the baseboard [S]; ConnectX-8 supports NDR 400 Gb/s [S]; HPE part number confirmed on the QuickSpecs at first build [A]

© 2026 Symbeeo. All rights reserved. · Public

Role	Product (exact model)	Qty	Capacity used	Why this unit
Storage	NetApp AFF A1K HA pair	1	2 PB usable of 14.688 PB raw ceiling [S]	One pair covers R8 with room for R14 [S]
Storage network	200 GbE ports on the A1K, 24 available per pair	8 used	2 per node [A]	Highest Ethernet speed listed on the A1K [S]
Scheduler	Slurm with fair-share	1	6 accounts	Open-source; not a vendor on the list [A]

4 · Design by subsystem

4.1 GPU nodes.

HPE ProLiant Compute XD685 in the 5U direct-liquid-cooled chassis, eight NVIDIA HGX B300 GPUs, two 5th Gen AMD EPYC processors, 24 DDR5-6400 RDIMMs, HPE iLO 6 [S]. Eight nodes per rack in the DLC form [S]; this design fills half a rack. Rejected: XD685 air-cooled at 6U — air cooling is listed for H200 only, not for B300 [S].

4.2 Compute fabric.

Two NVIDIA QM9700 switches, 64 ports of NDR 400 Gb/s each on 32 OSFP cages (two ports per cage), 51.2 Tb/s, onboard subnet manager [S]. Each node presents eight 400 Gb/s ports, one per GPU, on its ConnectX-8 adapters, split four to each switch [A]; the two switches are two rail planes with no links between them, so there is no spine tier [A]. Port arithmetic [T] derived: 4 nodes × 8 GPUs × 1 port = 32 links; 32 ÷ 2 switches = 16 ports per switch on day one (8 of 32 cages), 16 of 64 in use. R14's four more nodes bring it to 32 of 64 per switch; the pair tops out at 16 nodes (64 ports ÷ 4 per node). Optics: 32 node-side 400 Gb/s modules and, at two ports per cage, 16 switch-side twin-port modules (8 per switch); part numbers from NVIDIA's cable list at first build [A]. Rejected: Quantum-X800 Q3200-RA (72 XDR ports over 36 OSFP cages) [S] — R5 sets 400 Gb/s per GPU [T]; ConnectX-8 can run XDR at 800 Gb/s [S], so the Q3200-RA is the upgrade path if R5 is raised (32 links fit its 72 ports).

4.3 Storage.

One NetApp AFF A1K HA pair, 2U per controller, 14.688 PB raw ceiling and 2,048 GB memory per pair [S]; populated to 2 PB usable [T]. NFS for home and shared datasets; NVMe/TCP for scratch [A]. The A-Series scales to 12 HA pairs and 176.3 PB raw under NAS [S], so R14 growth adds shelves, not a new platform.

4.4 Storage network.

200 GbE, two ports per node into the A1K's 24 available 200 Gbps ports [S]; storage traffic kept off the InfiniBand fabric [A]. Rejected: 100 GbE — a single node at 8 GPUs can outrun 2 × 100 GbE on checkpoint writes [A].

4.5 Scheduler and fair-share.

Slurm with one account per group, fair-share weights set by the site's allocation committee, and a per-job GPU limit of 16 to keep any single job to at most half the cluster [A]. Usage reports per account monthly [A].

4.6 Management and out-of-band.

HPE iLO 6 on every node [S]; a management switch and console on their own VLAN, reachable before the fabric exists [A].

© 2026 Symbeeo. All rights reserved. · Public

5 · Failure behavior

Table 5. Failure behavior

Failure	Behavior	Source
One GPU fails	node drains in Slurm; 24 GPUs remain schedulable until repair	[A]
One node fails	8 of 32 GPUs lost; queued jobs reschedule; 16-GPU jobs still run	[A]
One QM9700 fails	16 of 32 node links lost, four per node; single-node jobs unaffected; multi-node jobs lose that rail plane until repair	[S] port count; [A] behavior
One A1K controller fails	HA partner takes over; no data loss	[S] HA pair; [A] takeover time
Facility water loss	all four nodes shut down on thermal limit	[A]
Slurm controller fails	running jobs continue; no new jobs until restart	[A]

6 · Limits

- The adapter is ConnectX-8, from NVIDIA's HGX B300 reference [S] and HPE's XDR adapter QuickSpecs [S*]; the XD685 QuickSpecs page could not be opened on Sep 27, 2026, so the HPE part number is confirmed at first build.
- The A1K page lists 200 Gbps as its top Ethernet speed and no 400 GbE; if the storage network must be 400 GbE, the platform choice changes.
- 2 PB usable is a target; effective capacity after ONTAP efficiency is set at first build.
- Fair-share is a policy; it needs a committee that owns the weights, or it decays.
- No burst to a hyperscaler in this design; a group that outgrows 32 GPUs waits.

7 · Lessons learned

None yet — this is a design study; populated after the first build against it.

8 · The author

WRITTEN BY

Diksha Shah

Principal AI Infrastructure Architect

NVIDIA Certified Professional: AI Operations, NVIDIA Certified Associate: AI Infrastructure and Operations, Google Cloud Professional Cloud Architect, AWS Certified Data Engineer

Signed off by Edwin Ozer, Founder.

© 2026 Symbeeo. All rights reserved. · Public

SYMBEEO

Each vendor's tool shows you their option. We show you all of them against one requirement.

Compared. Then Delivered.

hello@symbeeo.com · symbeeo.com

© 2026 Symbeeo. All rights reserved.

Product and company names and logos are their owners' trademarks, shown only to identify the lines Symbeeo sources. Naming a vendor or showing its logo does not mean the vendor sponsors or approves Symbeeo's comparisons or any recommendation Symbeeo makes.

Figures are as published by their sources on the dates given in this document; figures marked [A] are assumptions.

AI-003 · Version 1.3 · Oct 10, 2026