

JouleFlex — Compute price quote

Price documentation for §6.3, call 34/2026 (subcontractors / עבודות חוץ). Cloud GPU measurement runs and BMC-metered dedicated servers. Figures retrieved **29.08.2026**; verbatim vendor-published rates.

1. AWS EC2 GPU instances — us-east-1 (N. Virginia), Linux

Instance	GPU (count × model)	On-demand \$/hr	Spot \$/hr	On-demand \$/GPU-hr
g4dn.xlarge	1 × NVIDIA T4	0.526	0.220	0.526
g5.xlarge	1 × NVIDIA A10G	1.006	0.501	1.006
g6.xlarge	1 × NVIDIA L4	0.805	0.105	0.805
g6e.xlarge	1 × NVIDIA L40S	1.861	console	1.861
p3.2xlarge	1 × NVIDIA V100	3.060	1.381	3.060
p4d.24xlarge	8 × NVIDIA A100 40GB	21.958	13.067	2.745
p5.48xlarge	8 × NVIDIA H100	55.040	57.762	6.880

Sources (AWS-published, public): on-demand — pricing endpoint behind aws.amazon.com/ec2/pricing/on-demand/ (b0.p.awsstatic.com/pricing/2.0/.../US East (N. Virginia)/Linux/index.json); spot — public spot feed spot-price.s3.amazonaws.com/spot.js (aws.amazon.com/ec2/spot/pricing/). Per-GPU-hr = instance price ÷ GPU count. Spot prices vary continuously; values are the feed at retrieval. Full-page prints: [aws-ec2-on-demand-2026-08-29.pdf](#), [aws-ec2-spot-2026-08-29.pdf](#).

2. GPU Mart bare-metal dedicated servers — USA (dedicated IPv4 + IPMI)

Bare-metal servers with out-of-band **IPMI/BMC**, enabling PSU-input (wall) power reads via DCMI / Redfish — the socket-level measurement tier that GPU (NVML) and node (RAPL) telemetry cannot see. Monthly rate; rented short-term for the WP4 flexibility demonstration and cross-tier validation.

Plan	GPU	GPU memory	Price / month (USD)
Advanced Dedicated	1 × NVIDIA V100	16 GB HBM2	229.00
Enterprise Dedicated	1 × NVIDIA A100	40 GB HBM2	359.55
Enterprise Dedicated	1 × NVIDIA A100	80 GB HBM2e	1,559.00
Enterprise Dedicated	1 × NVIDIA H100	80 GB HBM2e	2,099.00

Source (public): gpu-mart.com/gpu-dedicated-server — all plans list “Dedicated IPv4 + IPMI”, USA datacenter, 99.9% uptime SLA. Retrieved 29.08.2026. Full-page print: [gpu-mart-dedicated-servers-2026-08-29.pdf](#). Equivalent bare-metal-with-BMC providers: Hivelocity, Latitude.sh, OpenMetal.

3. How this maps to the budget

Component	Plan	Planned (USD/yr)
AWS GPU-hours (~5,550/yr across families; spot for sweeps, on-demand where power control needs it)	\$1 above	≈ 16,060
Host CPU, storage, egress (allocation)	AWS	≈ 1,000
BMC-metered dedicated servers (short-term, WP4 + validation)	\$2 above	≈ 2,600
Planned total		≈ 19,660

Requested subcontractor line: **74,000 ₪/yr** (≈ \$23,125 at a conservative 3.20 ₪/\$; spot ≈ 2.96). The ≈ \$3,460/yr (15%) headroom covers spot-price movement, FX drift, and rate changes between submission and run; payment is against invoices (\$6.4), overhead-exempt (\$7.1).

מחירונים פומביים מתוארכים של AWS ו-GPU Mart. התדפיסים המלאים שמורים בתיקיית **/quotes** ומצורפים כהצעת המחיר לפי §6.3.