BARE METAL versus HYPERSCALER

A transparent monthly TCO model for Kubernetes infrastructure

Reference workload

344 vCPU | 688 GB RAM | 52 TB NVMe / SSD | 25 TB monthly egress

Central finding

The supplied model yields an estimated monthly run-rate of €26,034 for the AWS architecture and €10,469 for the KubeSailor bare-metal architecture, a modeled saving of €15,565 per month or approximately 60% on the monthly run-rate. The annualized difference is €186,780.

Technical economic note - version 1.0 - August 2026

Prepared as a decision-support document for engineering leaders, CTOs, finance teams and infrastructure architects.

Abstract

This technical economic note develops and audits a simple total-cost-of-ownership model for comparing a Kubernetes platform delivered on AWS against an equivalent bare-metal platform operated with KubeSailor. The reference demand vector is intentionally concrete: 344 vCPU, 688 GB of system memory, 52 TB of SSD/NVMe capacity and 25 TB of monthly egress. The objective is not to claim that cloud economics can be reduced to four retail prices. Rather, it is to show how a compact affine model can be constructed from published cloud price signals, a reference hardware cost pool and explicit architecture overheads, and then used consistently in a public-facing calculator.

Executive summary

- The supplied AWS formula is an architecture-loaded model: €42 per vCPU-month, €4.50 per GB-month, €120 per TB-month and €90 per TB of egress. At the reference point these terms sum to €26,034/month.
- The supplied bare-metal formula is a normalized cost-allocation model: €16 per vCPU-month, €1.70 per GB-month, €35 per TB-month, €15 per TB of egress and €1,600/month of fixed platform cost. These terms sum to €10,469/month.
- The modeled difference is €15,565/month, €186,780/year, with AWS at roughly 2.49x the bare-metal monthly run-rate in this reference case.
- The most important methodological point is that the unit coefficients are not intended to be mistaken for literal AWS invoice line items or isolated hardware purchase prices. They are demand-driver rates used to approximate a full architecture TCO.
- The model is most credible when the coefficient definitions are frozen, the underlying assumptions are published, and customers can replace them with region-specific quotes, actual utilization and their own operating-cost data.

Scientific status

This document is an engineering-economic model, not peer-reviewed academic research. Its value is reproducibility: every coefficient can be traced to a stated cost pool or pricing reference, recalibrated when inputs change, and stress-tested before procurement.

Keywords

Total cost of ownership; cloud economics; bare metal; Amazon EKS; Amazon EC2; EBS; data transfer; Kubernetes; capacity planning; cost allocation; FinOps.

1. Cost framework and research design

A fair comparison must place the two deployment models on the same economic boundary. The resource vector is therefore held constant, while the mechanism used to deliver the capacity differs. For AWS, the comparison includes compute, persistent block storage, Kubernetes control-plane charges and egress-related network cost. For bare metal, the comparison includes amortized infrastructure, power and cooling, networking, operations and a fixed platform layer.

1.1 Demand vector

Driver	Reference value	Unit	Role in the model
Compute	344	vCPU	Primary compute capacity
Memory	688	GB	System RAM capacity
Storage	52	TB	Persistent SSD/NVMe capacity
Egress	25	TB/month	Internet-bound data transfer

1.2 Model form

$$C\_month = F + r\_vCPU \cdot V + r\_RAM \cdot R + r\_storage \cdot S + r\_egress \cdot E$$

This affine form is deliberately compact. A detailed TCO model could include dozens of additional terms, such as IP addresses, load balancers, NAT gateways, snapshots, backup retention, observability, support contracts, spare hardware, cross-zone traffic and personnel. The coefficient model compresses these secondary effects into explicit loaded rates or a fixed platform term. AWS itself recommends cost modeling based on workload metrics and selecting the appropriate granularity rather than relying on a single generic spend number [6][7].

1.3 Why a coefficient model is defensible

- It creates a common unit of analysis across providers and deployment models.
- It avoids the false precision of pretending a heterogeneous architecture has a single invoice price.
- It supports what-if analysis: a user can change vCPU, RAM, storage and egress without rebuilding a full cloud bill.
- It makes the economics auditable. Each rate has a documented meaning, source and calibration method.

Boundary condition

The calculator should describe itself as an "indicative monthly run-rate" rather than a quotation. A production TCO audit should replace the generic coefficients with actual region, contract, hardware, utilization and operational data.

2. AWS reference architecture and observable price signals

To make the AWS side concrete, the reference architecture can be represented with Amazon EC2 compute-optimized instances that preserve the demand vector ratio. AWS documents the C7i family at 4 vCPU / 8 GiB for c7i.xlarge and 2 GiB of memory per vCPU, which matches the 344 vCPU / 688 GB requirement exactly at the capacity-ratio level [2]. A current secondary pricing snapshot lists c7i.xlarge in eu-west-1 at $0.1915/hour on demand [3].

2.1 A transparent AWS baseline

Component	Reference calculation	Monthly USD
EC2 compute	86 × c7i.xlarge × $0.1915/h × 730 h	$12,022.37
EBS gp3 storage	52,000 GB × $0.08/GB-month	$4,160.00
Internet egress	first 100 GB free; 10 TB at $0.09/GB; next 14.9 TB at $0.085/GB	$2,166.50
Amazon EKS	3 clusters × $0.10/h × 730 h	$219.00
Observable baseline total	Sum of above	$18,567.87

Using an illustrative FX translation of 1 EUR = 1.1380 USD (ECB reference rate dated 29 July 2026), this observable baseline is approximately €16,316/month [8]. The important result is not the exact FX conversion; it is the gap between the observable component prices and the higher loaded rate used by the public calculator.

2.2 Why the real AWS bill is higher than a four-line price list

- EKS bills per cluster-hour. Standard support is currently $0.10 per cluster-hour, while extended support is $0.60 per cluster-hour [1].
- EBS gp3 is priced on provisioned storage, with additional charges possible for provisioned IOPS and throughput above the baseline [4].
- EC2 data transfer pricing is tiered and the free 100 GB internet egress allowance is aggregated; transfer within AWS can also incur separate charges depending on path and availability-zone topology [5].
- A real Kubernetes platform commonly introduces load balancers, public IPv4, logging/monitoring, snapshots, NAT or gateway services, multi-AZ redundancy, headroom and non-productive capacity.

Interpretation

The calculator coefficient should therefore be called a loaded AWS rate, not an AWS list price. Its purpose is to approximate the architecture-level monthly cost from four demand drivers, while the underlying AWS services remain priced according to their own meters.

3. Deriving and defending the AWS coefficients

The supplied AWS coefficients are €42/vCPU-month, €4.50/GB-month, €120/TB-month and €90/TB egress. They reproduce the screenshot exactly, but a defensible methodology must explain what they stand for. The correct interpretation is a normalized architecture-loaded rate: the coefficient captures the direct service cost plus the expected share of platform overhead associated with that demand driver.

3.1 Compute coefficient: €42 per vCPU-month

The c7i.xlarge reference gives a public list-price signal of $0.1915/hour in eu-west-1 [3]. At 730 hours/month, that is $139.80 per instance-month, or $34.95 per vCPU-month. At the illustrative 1.1380 USD/EUR rate, the direct compute signal is about €30.71 per vCPU-month. The calculator coefficient of €42 is therefore approximately 1.37 times the direct compute signal.

That uplift is intentionally interpreted as an architecture-loading factor for Kubernetes capacity: scheduling headroom, multi-AZ placement, operational spare capacity, non-ideal packing, platform components and the fact that a real deployment is not a pure EC2-only workload. The coefficient is therefore conservative rather than a statement that AWS literally invoices €42 per vCPU.

3.2 Storage coefficient: €120 per TB-month

AWS publishes gp3 storage at $0.08/GB-month in regions using that reference price, with performance charges above the included 3,000 IOPS and 125 MiB/s baseline [4]. A pure 52 TB gp3 storage line at $0.08/GB-month is therefore far below €120/TB-month after FX conversion. The higher coefficient is best understood as a storage-platform rate that allows for production storage overheads: performance provisioning, snapshots/backups, replica or spare capacity, and the fact that a calculator should not assume every customer will use the absolute lowest storage configuration.

3.3 Egress coefficient: €90 per TB

AWS internet egress pricing is tiered. The first 100 GB/month is currently free, followed by published tiers beginning at $0.09/GB for the first 10 TB/month and $0.085/GB for the next 40 TB/month in the general EC2 pricing schedule [5]. At 25 TB/month, the weighted reference cost is approximately $2,166.50 before other network-path charges. The calculator's €90/TB rate is deliberately close to this order of magnitude after FX, while leaving room for traffic-path complexity and pricing drift.

3.4 RAM coefficient: €4.50 per GB-month

RAM is not separately metered on ordinary EC2 instances; it is bundled into instance types. Consequently, the €4.50/GB-month term is an allocation rate, not a retail AWS memory price. It exists to let the simplified calculator respond sensibly to changes in memory-heavy workloads while the reference architecture maintains a fixed CPU-to-memory ratio. This is a modeling convenience and should be documented as such.

Defence in one sentence

The AWS coefficients are defensible because they are not presented as four independent AWS invoices; they are calibrated, architecture-loaded demand-driver rates anchored to public AWS price signals and inflated only where additional production overhead is expected.

4. Bare-metal reference architecture and cost pool

The bare-metal side asks what it costs to own or lease enough physical infrastructure to deliver the same capacity continuously. The monthly economic cost is a pool of hardware lifecycle, power/cooling, rack and network, support/spares, and platform operations.

4.1 Recommended accounting boundary

Cost pool	What it contains	How it enters the model
Compute pool	CPU-capable server capacity; lifecycle and support	Allocated through vCPU rate
Memory pool	RAM capacity and the incremental server mix needed to provide it	Allocated through GB rate
Storage pool	Enterprise NVMe/SSD, controllers, endurance and replacement reserve	Allocated through TB rate
Network pool	Transit, upstream connectivity and expected egress	Allocated through TB egress rate
Platform pool	Kubernetes control-plane services, monitoring, registry, backup, automation, rack and shared operations	Fixed monthly term

4.2 Why a bare-metal cost is not just the purchase price

- Hardware is amortized or lease-equivalent, then increased by power, cooling, rack/network, support and spare-capacity costs.
- HA and hardware-failure policies require reserve capacity above the exact steady-state demand vector.
- A private Kubernetes platform still carries monitoring, backups, upgrades, security and operator effort even when the software is open source.

Why KubeSailor changes the equation

KubeSailor's role is to reduce the operational friction normally associated with private infrastructure by packaging cluster construction, automation and platform components. The economic case therefore depends not only on hardware prices but also on how much platform-engineering effort is avoided.

4.3 Coefficients used by the calculator

Driver	Rate	Interpretation
vCPU	€16 / vCPU-month	Allocated compute cost pool
RAM	€1.70 / GB-month	Allocated memory cost pool
NVMe	€35 / TB-month	Allocated local-storage cost pool
Egress	€15 / TB	Network/transport cost allocation
Platform	€1,600 / month	Shared, demand-insensitive infrastructure and operations

5. Deriving and defending the bare-metal coefficients

The supplied bare-metal coefficients are most appropriately derived using activity-based costing. Begin with the monthly cost of the actual reference infrastructure pool. Partition that cost into compute, memory, storage, networking and shared platform pools. Then divide each pool by the baseline demand capacity that the pool is intended to support.

Rate_i = Allocated monthly cost pool_i / Baseline demand_i

5.1 Demonstration for the reference case

Driver	Baseline	Rate	Monthly allocation
vCPU	344	€16	€5,504.00
RAM	688 GB	€1.70/GB	€1,169.60
NVMe	52 TB	€35/TB	€1,820.00
Egress	25 TB	€15/TB	€375.00
Platform	Fixed	€1,600	€1,600.00
Total	-	-	€10,468.60

This calculation is transparent, but the rates must be calibrated from a real hardware and operations dataset before commercial publication. In a customer-facing calculator, a rigorous KubeSailor implementation would ideally maintain a versioned rate card: hardware purchase or lease quotes, depreciation period, expected utilization, power tariff, network contracts, staffing assumptions and spare-capacity policy.

5.2 How to build the underlying cost pool in practice

- Collect three or more current server quotations covering CPU, RAM and enterprise NVMe capacity.
- Choose a depreciation or lease term, such as 36 or 48 months, and document the residual-value assumption.
- Estimate steady-state power and cooling from measured or vendor-rated consumption; apply the local electricity tariff.
- Add rack space, switching, upstream bandwidth, maintenance/support, hardware spares and a platform-operations budget.
- Set a target utilization or reserve policy. For example, if 20% capacity is held back for resilience, the allocated demand base should not exceed 80% of installed capacity.
- Recalculate the coefficients whenever major hardware, energy, bandwidth or staffing assumptions change.

Important caveat

The rates €16/vCPU, €1.70/GB and €35/TB are therefore defendable as cost-allocation coefficients, not as universal market prices for bare metal. Their scientific strength is reproducibility and calibration, not universality.

6. Worked reference case: reproducing the calculator

The screenshot supplied with this study is reproduced by the following deterministic calculation. No rounding is applied until the monthly totals are produced.

6.1 AWS result

$$\text{AWS} = 344\cdot 42 + 688\cdot 4.5 + 52\cdot 120 + 25\cdot 90$$

Term	Calculation	EUR/month
vCPU	344 × €42	€14,448
RAM	688 × €4.50	€3,096
Storage	52 × €120	€6,240
Egress	25 × €90	€2,250
AWS total	Sum	€26,034

6.2 KubeSailor result

$$\text{KubeSailor} = 344\cdot 16 + 688\cdot 1.7 + 52\cdot 35 + 25\cdot 15 + 1600$$

Term	Calculation	EUR/month
vCPU	344 × €16	€5,504
RAM	688 × €1.70	€1,169.60
Storage	52 × €35	€1,820
Egress	25 × €15	€375
Platform	Fixed	€1,600
KubeSailor total	Sum	€10,468.60 → €10,469

6.3 Savings

$$\text{Monthly saving} = €26{,}034 - €10{,}469 = €15{,}565$$

$$\text{Annual saving} = €15{,}565 \times 12 = €186{,}780$$

$$\text{Savings rate} = €15{,}565 / €26{,}034 = 59.8\% \approx 60\%$$

The screenshot's result is therefore mathematically consistent with the supplied coefficients. The model should preserve this exact transparency: every displayed value should be reproducible from the four sliders and the published rate card.

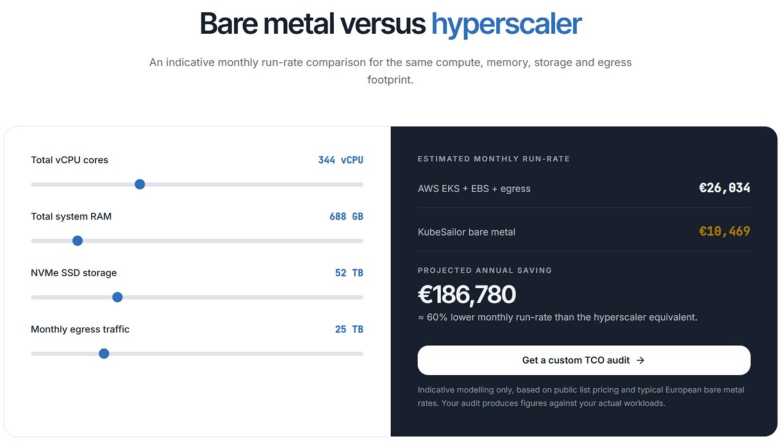

Figure 1. Supplied calculator output for the reference case. The values shown in the interface are reproduced by the formulas in this section.

7. Model validation and sensitivity

A useful economic model should make uncertainty visible. In the reference case, the vCPU and storage terms dominate both sides of the comparison. That means procurement decisions are most sensitive to hardware utilization, compute pricing, storage design and redundancy assumptions.

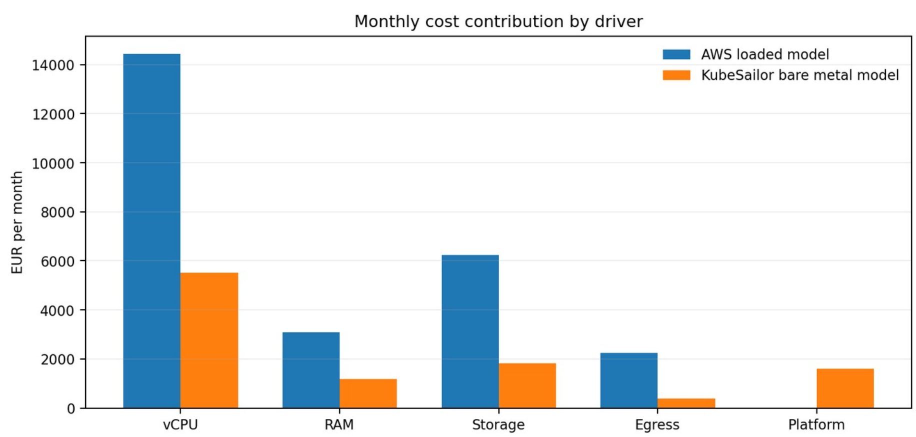

Figure 2. Contribution of each demand driver to the monthly totals.

7.1 Component shares in the reference case

Component	AWS	KubeSailor
Compute	55.5%	52.6%
Memory	11.9%	11.2%
Storage	24.0%	17.4%
Egress	8.6%	3.6%
Fixed platform	0.0%	15.3%

The fixed platform term is strategically important: it means the bare-metal model becomes comparatively more efficient as the platform is loaded with more capacity, while very small deployments may not justify the fixed infrastructure and operational burden.

8. Break-even interpretation and decision thresholds

The reference model yields AWS = €26,034 and bare metal = €10,469. If all AWS loaded rates were scaled by a common factor while the bare-metal model remained unchanged, AWS would need to fall to approximately 40.2% of the loaded baseline to reach cost parity. Equivalently, the current model contains a 59.8% cost gap.

Break-even AWS factor = €10,468.60 / €26,034 = 0.402

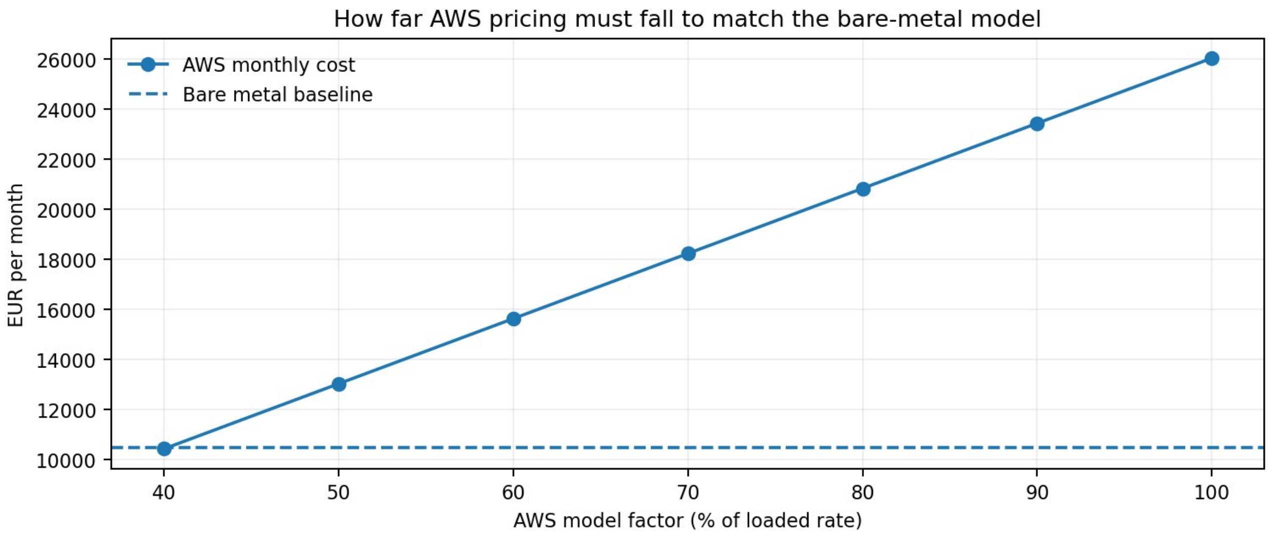

Figure 3. Common-factor sensitivity of the AWS loaded model. The chart is illustrative; real AWS discounts are not uniform across all services.

8.1 What could move the break-even point?

- AWS commitments such as Savings Plans or Reserved pricing can materially reduce compute cost relative to on-demand, moving the break-even point toward AWS.
- Better EC2 right-sizing and lower storage performance requirements reduce the loaded AWS rate.
- On bare metal, lower utilization, higher staffing, expensive connectivity, shorter hardware lifecycles or high redundancy targets increase the coefficient set.
- Bursting and elastic demand can favor AWS because the cloud lets customers avoid paying for idle physical capacity.
- Conversely, stable 24×7 workloads with high resource utilization can favor owned or leased hardware because the capital cost is spread across continuous usage.

A practical rule

The economic case becomes strongest for bare metal when demand is large, stable, highly utilized and performance/storage intensive, while the economic case for the hyperscaler strengthens when demand is volatile, short-lived or difficult to forecast.

9. Limitations, omissions and responsible use

The calculator is intentionally small. A professional TCO audit should not hide the fact that several variables can be materially important. The public-facing article should therefore state both what the calculator includes and what it does not.

9.1 Included in the simplified comparison

- Recurring compute capacity, memory, storage and internet egress.
- A Kubernetes platform charge on the AWS side through the loaded coefficients and an explicit fixed platform term on bare metal.
- A consistent monthly time basis and annualization of the modeled savings.

9.2 Not explicitly modeled

- Application licensing; commercial databases; proprietary observability or security products.
- Detailed load-balancer, NAT gateway, public IPv4, snapshots, backup retention, object storage and inter-region traffic charges.
- Staffing rates and opportunity cost of platform engineering unless embedded in the bare-metal platform pool.
- Financing cost, tax treatment, VAT, currency volatility, depreciation accounting method and residual hardware value.
- Disaster recovery requirements, secondary sites, geographic redundancy and business continuity requirements.

9.3 Responsible commercial wording

The safest customer-facing statement is: "This is indicative modeling based on public list pricing, a reference architecture and normalized bare-metal cost drivers. Actual TCO varies by AWS region, discounts, workload behavior, availability requirements, hardware quotes, energy prices and operating model. A custom audit replaces the defaults with measured or quoted inputs."

Recommended website language

Do not promise "60% cheaper than AWS" as an unconditional claim. Use "up to" only when supported by a defined benchmark population. For this reference case, the precise statement is "approximately 60% lower modeled monthly run-rate under the stated assumptions."

9.4 Decision checklist for a CTO

Question	Why it matters
Is demand stable 24×7?	Favors amortized physical capacity when utilization is high.
Can the workload burst unpredictably?	Favors elasticity and consumption pricing.
What is the required HA/DR topology?	Determines how much extra capacity each model needs.
What are actual hardware and energy quotes?	Calibrates the bare-metal coefficients.
What AWS discounts are available?	Calibrates the hyperscaler coefficients.
Who operates the platform?	Converts engineering effort into a real cost line.

10. Conclusions and reference sources

For the specified infrastructure demand of 344 vCPU, 688 GB RAM, 52 TB storage and 25 TB monthly egress, the supplied calculator produces a monthly AWS run-rate of €26,034 and a KubeSailor bare-metal run-rate of €10,469. The resulting annualized difference is €186,780, corresponding to approximately 60% lower modeled monthly cost for the bare-metal option.

The coefficients are defendable when they are described correctly: AWS rates are architecture-loaded demand-driver rates anchored to public EC2, EBS, EKS and network pricing; bare-metal rates are activity-based cost allocations derived from an amortized infrastructure pool and an explicit fixed platform cost. They should not be presented as universal retail prices. Their value is that they are simple, auditable and replaceable.

Final recommendation

Keep the four-slider calculator. Publish the coefficient definitions, methodology and assumptions next to it. Version the rate card. Refresh cloud price anchors periodically. Offer a custom TCO audit for prospects whose architecture differs materially from the reference case. This turns the calculator from a marketing graphic into a traceable economic model that a CTO, architect or finance partner can challenge and reproduce.

References

[1] Amazon Web Services. Amazon EKS Pricing. Standard support: $0.10 per cluster-hour; extended support: $0.60 per cluster-hour. https://aws.amazon.com/eks/pricing/

[2] Amazon Web Services. Amazon EC2 Compute Optimized Instances / C7i. c7i.xlarge: 4 vCPU, 8 GiB, EBS-only. https://aws.amazon.com/ec2/instance-types/c7i/

[3] DoiT Compute. AWS c7i.xlarge, eu-west-1. Current on-demand snapshot used as a public pricing cross-check: $0.1915/hour. https://www.doit.com/compute/spot/eu-west-1/c7i.xlarge

[4] Amazon Web Services. Amazon EBS Pricing. gp3 storage reference: $0.08/GB-month; additional IOPS/throughput are separately priced above baseline. https://aws.amazon.com/ebs/pricing/

[5] Amazon Web Services. EC2 On-Demand Instance Pricing / Data Transfer. 100 GB/month free and tiered internet egress rates beginning at $0.09/GB. https://aws.amazon.com/ec2/pricing/on-demand/

[6] Amazon Web Services. Cloud Financial Management - Cost Optimization Pillar. Emphasizes cost transparency, planning, unit metrics and ongoing optimization. https://docs.aws.amazon.com/wellarchitected/latest/cost-optimization-pillar/practice-cloud-financial-management.html

[7] Amazon Web Services. COST06-BP01 Perform cost modeling. Recommends workload-specific cost modeling and validation using the selected resource types and usage patterns. https://docs.aws.amazon.com/wellarchitected/latest/framework/cost_type_size_number_resources_cost_modeling.html

[8] European Central Bank. Euro foreign exchange reference rates. Reference rate used for the illustrative USD/EUR translation: 1 EUR = 1.1380 USD on 29 July 2026. https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html

Pricing is time-sensitive. Verify current AWS regional list prices and any commercial discounts before using this document for procurement or contractual commitments.