KloudMate

Observability Cost Benchmark and Control Guide

2026

What telemetry really costs, why bills drift, and how engineering teams can regain control

First edition · September 2026 · Research snapshot 17 September 2026

GUIDE

Contents

Executive summary 04

1 The bill is a pipeline, not a SKU 07

2 The 2026 pricing map 10

3 Three normalized workload illustrations 16

4 Why costs rise 20

5 A procurement benchmark that survives scrutiny 24

6 A 90-day cost-control program 26

7 What we believe at KloudMate 29

8 Outlook for the rest of 2026 30

9 Methodology and limitations 31

10 Sources 32

ABOUT THIS GUIDE

A benchmark and control guide, not a survey-based “state of the industry” report.

It is built from public vendor pricing, vendor documentation, and published third-party surveys, not from original fieldwork. Where a number comes from a vendor that sells cost optimization, we say so. Where we compare platforms, we show the metering boundary instead of pretending the units match.

KloudMate sells an observability platform, and our own pricing appears in the benchmark table. We applied the same normalization tests to ourselves as to everyone else, but treat our inclusion the way you would treat any vendor-authored comparison: as a starting point, not a verdict.

Who this is for

- Founders and engineering leaders who own an observability bill that grew faster than expected.
- SRE and platform teams asked to cut telemetry spend without losing production answers.
- FinOps practitioners extending cloud cost discipline to observability.
- Anyone evaluating vendors who wants a comparison method that survives procurement scrutiny.

A future companion piece, the State of Observability Costs 2026, will be built on original survey and billing research.

EXECUTIVE SUMMARY

Observability has a value problem before it has a price problem

Teams collect more telemetry than they can use, buy overlapping tools, and discover the economics only after production traffic, cardinality, and retention rise.

74%

said cost was a top tool-selection priority, more than ease of use or interoperability

Grafana Labs 2025 · n=1,255 [1]

13%

of collected telemetry was actively used on average; 84% of companies used less than a quarter

Sawmills 2025 · vendor-sponsored [2]

8

observability technologies per organization on average; 39% called complexity and overhead their largest obstacle

Grafana Labs 2025 [1]

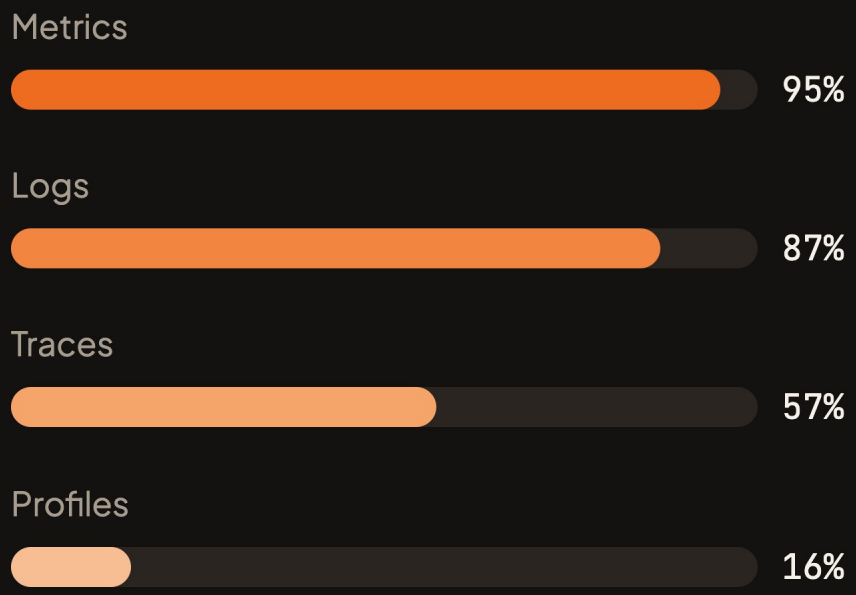
WHAT THE EVIDENCE SHOWS

Public pricing has fragmented, and OpenTelemetry does not reduce cost by itself.

In 2026 the market uses host-based, ingest-based, event-based, active-series, sample-based, stored-GB, query-compute, and hybrid plans. A low headline rate can cover a different unit, and a different part of the data lifecycle, than a competitor's rate.

Sending the same high-volume, high-cardinality data through an OpenTelemetry Collector makes expensive data portable. It does not make it cheaper.

Signals in use, across 1,255 respondents



71% used Prometheus and OpenTelemetry in some capacity.

Source: Grafana Labs Observability Survey 2025 [1]

• OUR CONCLUSION

The cheapest telemetry is not the telemetry with the lowest per-GB vendor price. It is the smallest set of signals that can still answer production questions, with clear ownership, retention, and unit economics.

SECTION 01

The bill is a pipeline, not a SKU

The five stages of a telemetry bill

Where each charge actually lands

Why two $0.40/GB offers differ

THE COST MODEL

```
Total observability cost = collection + network + processing + ingest + indexing + storage/retention
+ query/rehydration + seats/hosts + support + engineering labor + incident risk
```

Most pricing pages show only some of these terms. Map every charge to five stages.

Generate

Application log lines, spans, metric samples, profiles, RUM events, synthetic runs.

Move

Agents, collectors, egress, cross-region or cross-cloud transfer.

Process

Parsing, enrichment, redaction, aggregation, sampling, routing.

Store

Hot index, object archive, retained series and events, replicas.

Use

Queries, scanned bytes, rehydration, dashboards, alerts, users, AI tokens or executions.

THE METERING BOUNDARY

This is why two “$0.40/GB” offers can produce different bills.

One may meter uncompressed ingress, another stored data after processing, another processing plus a separate write or query charge.

A useful cost model runs across all five stages: generate, move, process, store, use. Most pricing pages show only some of those terms, so the comparison has to be rebuilt stage by stage rather than read off the headline rate.

Same headline rate, three different meters

Vendor A meters uncompressed ingress

Generate	**Move**	Process	Store	Use

Vendor B meters stored data after processing

Generate	Move	Process	**Store**	Use

Vendor C meters processing, then charges writes and queries

Generate	Move	**Process**	Store	**Use**

Never compare unit prices without the metering boundary.

SECTION 02

The 2026 pricing map

Nine platforms at public list price

What each headline number counts

The comparison trap

HOW TO READ THE MAP

A pricing-model map, not an apples-to-apples ranking.

It uses vendor-published public rates available on 17 September 2026. Every entry names the headline number and what the buyer must still resolve before the number means anything.

Prices were checked against vendor pages on the research date and will drift. Contract discounts, regional rates, taxes, support tiers, retention, and overages can change the result.

Eight metering models in the 2026 market

- Host-based
- Ingest-based
- Event-based
- Active-series
- Sample-based
- Stored-GB
- Query-compute
- Hybrid

Two offers can share a unit and still meter it at a different stage of the data lifecycle.

THE 2026 PRICING MAP

Public list prices · 17 September 2026

Datadog

PUBLIC PRICING SIGNAL

Infrastructure Pro from $15/host/month billed annually. APM from $31/host/month with Infrastructure attached, or $36 standalone. Log ingestion from $0.10/ingested GB; 15-day log indexing at $1.70 per million events. [3][4]

STILL TO RESOLVE

Host count, container allotments, indexed event count, retention tier, custom metrics, rehydration and outbound forwarding. Excess committed log volume is charged 50% above annual price.

Splunk Observability Cloud

PUBLIC PRICING SIGNAL

Infrastructure from $15/host/month; App & Infra $60/host/month; End-to-End $75/host/month, billed annually. Splunk APM alone from $55/host/month. [5]

STILL TO RESOLVE

Exact bundle, host definition, logs purchased separately, RUM, synthetic and database add-ons, and contract terms.

New Relic

PUBLIC PRICING SIGNAL

100 GB/month free, then original data at $0.40/GB or Data Plus at $0.60/GB. EU storage adds $0.05/GB. Core users start at $49/user; Full Platform entry pricing varies by edition and eligibility. [6]

STILL TO RESOLVE

User mix or compute model, retention, synthetics, advanced compute, support edition, and whether processing rules reduce saved ingest. New Relic says egress and hydration are free.

Grafana Cloud

PUBLIC PRICING SIGNAL

Pro platform fee $19/month. Metrics include 10,000 active series, then $6.50/1,000 active series at the first public tier. Logs include 50 GB/month. Traces: $0.05/GB processing and $0.40/GB writing after 50 GB, 30-day retention. [7]

STILL TO RESOLVE

DPM adjustments, log process/write/retain/query components, retention, query fair-use, and the $25,000/year enterprise minimum. Adaptive Metrics and Adaptive Logs savings are product claims.

Honeycomb

PUBLIC PRICING SIGNAL

Free up to 20M events/month and 100M metric data points. Pro from $150/month for 50M events, scaling to 750M. Telemetry collection from $0.10/GB. Seats and queries are unlimited. [8]

STILL TO RESOLVE

Event definition and event size, metrics allowance, trigger and SLO needs, enterprise volume, and whether Refinery processing is self-managed.

Elastic Observability Serverless

PUBLIC PRICING SIGNAL

Logs Essentials ingest as low as $0.07/GB, retention as low as $0.017/GB-month. Complete lists $0.09/GB ingest for logs and traces, $0.023/GB for metrics, retention $0.019 and $0.005/GB-month. Egress: 50 GB free, then $0.05/GB. [9]

STILL TO RESOLVE

“As low as” volume assumptions, retained compressed size, region, query workload, support (5% to 15% above Standard), and add-ons.

AWS CloudWatch

PUBLIC PRICING SIGNAL

No minimum. New OpenTelemetry metrics are $0.50 per uncompressed GB ingested, including 15 months of storage. The free tier includes 5 GB for Logs categories and 10 custom or detailed metrics. Pricing varies by region. [10]

STILL TO RESOLVE

Region, log class, ingestion, archive, Insights scanned bytes, Live Tail, API calls, alarms, dashboards, cross-account copies, and AWS data transfer.

Coralogix

PUBLIC PRICING SIGNAL

Published pipelines include frequent-search logs at $1.15/GB and monitoring logs at $0.50/GB. One unit equals $1.50 of signal usage. [11]

STILL TO RESOLVE

Unit commitment, pipeline mix, daily quota and overage, BYO-S3 expense, hot retention, and regional or contract terms. Coralogix's compression and savings numbers are vendor claims.

KloudMate

PUBLIC PRICING SIGNAL

Free includes 10 GB logs, 10 GB traces, and 15M metric samples. Pro is $99/month used as a telemetry budget: logs and traces $0.40/GB, metrics $0.30/million samples, with no host or user fee. Public volume tiers show $0.32/GB and $0.225/million after stated thresholds. [12][13]

STILL TO RESOLVE

Retention multiplier, Enterprise custom rates, required feature tier, and actual signal mix. Default Pro retention is 14 days for logs and metrics, with traces up to one year.

THE COMPARISON TRAP

Do not read “Elastic at $0.09/GB” and “New Relic at $0.40/GB” as a 4.4× total-cost conclusion.

Elastic separately meters retention and may add support. New Relic includes product access under a hybrid model and describes ingest after its processing rules. Datadog’s logs mix GB ingestion with event indexing. Grafana splits processing and writing. Honeycomb charges around events. KloudMate meters metrics as samples.

What each headline number actually counts

Datadog	GB ingestion plus event-based indexing
New Relic	product access under a hybrid model, ingest counted after its processing rules
Elastic	ingest GB, with retention metered separately and support possibly added
Grafana Cloud	processing and writing split; metrics by active series
Honeycomb	events
KloudMate	GB for logs and traces; samples for metrics

A valid proof of concept replays the same telemetry, for the same period, with the same retention and query tests, then reconciles both invoices.

SECTION 03

Three normalized workload illustrations

Example A: 10 GB of logs per day

Example B: 100 GB of logs per day

Example C: metric cardinality

EXAMPLE A: 10 GB OF LOGS PER DAY

Monthly volume: 300 GB.

These are arithmetic examples, not quotes. They isolate a small slice of each platform and intentionally omit seats, hosts, support, queries, taxes, discounts, regional variation, and most features.

Datadog's searchable indexing is event-based. At an average event of 1 KB, 300 GB is roughly 300 million decimal events, subject to exact size, exclusions, commitment, and Datadog's counting rules. Average event size therefore changes the result sharply.

One slice of the bill, at public list price

$120

flat $0.40/GB ingest arithmetic. KloudMate Pro's $99 minimum is already consumed, so this slice is about $120 before any retention extension

$80

New Relic's data component beyond the published 100 GB allowance, before users, compute and add-ons

$27 + $5.70

Elastic Complete ingest at $0.09/GB, plus $0.019/GB-month if the retained billable footprint equalled the full 300 GB for one month

$30 + ~$510

Datadog ingestion at $0.10/GB, plus 15-day indexing at $1.70 per million events

EXAMPLE B: 100 GB OF LOGS PER DAY

Monthly volume: 3,000 GB.

The lesson is not that one vendor always wins. It is that GB/day alone is insufficient when another vendor bills events, hosts, active series, or storage.

Every figure here is simple multiplication on public list rates, shown with its inputs so a reader can rerun it. They exclude tax and do not constitute vendor quotes.

The same slice at ten times the volume

$1,200

flat $0.40/GB arithmetic

$1,160

New Relic original-data component after the 100 GB free allowance

$270

Elastic Complete ingest at $0.09/GB, plus retention, support, egress and add-ons as applicable

$300 + $5,100

Datadog ingestion at $0.10/GB, then 3 billion events at $1.70 per million for the indexing term

EXAMPLE C: METRIC CARDINALITY

For Prometheus-style metrics, series count is roughly metric names × unique combinations of label values.

A metric with 20 services, 10 regions, 5 status codes, and 50 customer-tier or tenant-like values can create 500,000 series before other labels.

Removing a single unbounded label can matter more than negotiating a small unit discount.

Where the multiplication happens

```
20 services × 10 regions × 5 status codes × 50
tenant-like values = 500,000 series
```

$3,185/mo

Grafana Cloud: 500,000 active series minus the included 10,000, at $6.50 per 1,000, before DPM adjustment and volume tiers

86.4 billion

KloudMate meters samples: one continuously active series at a 15-second interval emits about 172,800 samples/month, and one million samples cost $0.30 at list price

Metering model determines which behavior is expensive.

SECTION 04

Why costs rise

Logs: verbosity × traffic

Metrics: cardinality × frequency

Traces: request volume × span fan-out

Retention: daily ingest × time

Tool and team overlap

Logs

`verbosity × traffic`

COMMON CAUSES

- Debug logs left on in production
- Duplicate collection
- Multiline stack traces
- Health-check and access logs
- Large payloads
- High-cardinality fields in every event

CONTROLS

- Structured logging and a severity policy
- Source-side filtering
- Payload truncation and duplicate detection
- Tiered routing
- Short hot retention, object storage for audit data

A 2 KB event costs twice a 1 KB event under GB pricing, but the same under pure event pricing. The reverse can happen with many tiny events.

Metrics

`cardinality × frequency`

COMMON CAUSES

- User IDs, request IDs and pod UIDs as labels
- URLs and unconstrained tags
- One metric turned into millions of series
- Faster scrape intervals

CONTROLS

- Label allowlists
- Recording rules and aggregation
- Stale-series reports
- Lower frequency for low-value metrics
- Business telemetry separated from operational alerting

Faster scrape intervals multiply samples and data points, so how often you scrape is a pricing decision too.

Traces

`request volume × span fan-out`

THE TRADE-OFF

- Head sampling reduces volume cheaply but can miss rare failures
- Tail sampling keeps decisions informed by outcome, but needs buffering and collector capacity

RULES WORTH KEEPING

- Errors
- High latency
- Important tenants
- Novel routes
- A representative baseline

A request that crosses 20 services can create dozens or hundreds of spans. OpenTelemetry documents head and tail sampling as different trade-offs. [14]

Retention

`daily ingest × time`

THE UNEXAMINED DEFAULT

- Keeping everything hot for 30, 90, or 365 days

GIVE EACH A NAMED CLASS

- Operational troubleshooting
- SLO calculation
- Security investigation
- Statutory retention

These are different use cases with different time horizons. One retention setting for all of them keeps data hot that nobody needs hot.

TOOL AND TEAM OVERLAP

Grafana’s survey average of eight technologies is a warning.

The direct cost is duplicated licence and ingest. The larger cost is split context, duplicated dashboards, agents, alert rules, access controls, and expertise.

Self-hosted software avoids a SaaS invoice but adds compute, storage, upgrades, on-call load, capacity planning, backups, and specialist time.

Where the duplication lands

What you pay for twice

- Licences and ingest
- Dashboards and alert rules
- Agents and collectors
- Access controls
- Context and expertise

What self-hosting adds instead

- Compute and storage
- Upgrades and backups
- On-call load
- Capacity planning
- Specialist time

SECTIONS 05 AND 06

Benchmark and control

A 30-day shadow benchmark

Days 1–30: make the bill explainable

Days 31–60: reduce low-value volume

Days 61–90: make control continuous

01	**Freeze the workload**	Capture daily GB by signal, events, average event size, active series, samples/minute, spans, hosts, containers, users, query volume, and desired retention.
02	**Run both platforms in parallel**	Route the same production sample through OpenTelemetry where possible. Do not compare one vendor's sampled feed with another vendor's full feed.
03	**Test five investigations**	Deployment regression, slow endpoint, failed dependency, noisy host or pod, and a 30-day historical incident.
04	**Measure outcome and cost**	Ingestion success, query latency, time to diagnosis, alert quality, dropped data, engineer hours, and invoice by meter.
05	**Normalize annual cost**	Committed base, expected growth, 95th-percentile burst, support, retention, network, archive, migration, and operating labor.
06	**Stress the model**	Recalculate at 0.5×, 1×, 2×, and 5× traffic; double cardinality separately. A plan that is cheap at today's usage may have a steep next step.
07	**Publish the worksheet internally**	If nobody can reproduce the estimate from raw usage, the estimate is not ready for a contract.

DAYS 1-30

Make the bill explainable.

This phase does not change what you collect. It sets ownership, attribution and a baseline, so you can measure later cuts instead of guessing at them.

Until most telemetry is mapped to a team, a service, an environment, a signal and a retention class, a reduction target means very little.

Six moves

- Name one observability cost owner and one engineering owner per service.
- Build a daily unit-cost dashboard: cost per million requests, per active customer, per service, and as a percentage of cloud spend.
- Attribute at least 80% of telemetry to team, service, environment, signal, and retention class.
- Find duplicate agents and exporters, and the top 20 sources by spend.
- Set budget alerts at 50%, 75%, 90%, and forecasted overrun.
- Record the baseline: total spend, GB/day, active series, samples, events, spans, hot-storage days, and queries.

DAYS 31-60

Reduce low-value volume.

This is where collection changes. Every cut points at a named use case, and every exception carries an owner and an expiry date.

Short hot retention is only safe once the data is split into real-time, investigation, compliance and archive pipelines, and rehydration has been tested before anyone relies on the archive.

Six moves

- Remove debug, health-check, and duplicate logs in production unless they support a named use case.
- Apply metric label limits and delete unused recording rules.
- Introduce trace sampling that always keeps errors and latency outliers.
- Split data into real-time, investigation, compliance, and archive pipelines.
- Shorten hot retention; validate rehydration before relying on archive.
- Require every exception to have an owner and an expiry date.

DAYS 61-90

Make control continuous.

Cost control belongs in service reviews and deployment checks, not in a quarterly clean-up.

Suggested first-quarter targets: 90% cost allocation, zero unexplained week-over-week growth over 20%, 100% of high-volume sources with a retention class, and a measured 20% to 40% reduction in low-value ingest. The reduction target is an operating goal, not an industry guarantee.

Six moves

- Add telemetry budgets to service reviews and deployment checks.
- Detect week-over-week volume and cardinality regressions.
- Showback cost to teams; use chargeback only when attribution is trusted.
- Review unused dashboards, alerts, custom metrics, and seats quarterly.
- Reprice at contract anniversaries using measured workload, not vendor calculator defaults.
- Keep instrumentation portable and routing policy outside application code where practical.

WHAT WE BELIEVE AT KLOUDMATE

We wrote this guide and we sell an observability platform, so our incentives are on the table.

If buyers get better at benchmarking, vendors with honest pricing do better, and we would like to be one of them.

We do not think the market needs another claim that one platform is universally cheaper. Public list prices cannot prove that: workloads, contracts, and metering boundaries differ too much. KloudMate's public model is simple: logs and traces by GB, metrics by samples, no host or user fee. It will suit some workloads and not others.

Buyers deserve three things

01 A bill they can reproduce from their own usage.

02 Control before ingestion, not only a discount after waste has arrived.

03 Portability, so instrumentation decisions outlive a vendor contract.

Our own pricing belongs in the same benchmark, under the same traffic replay and invoice reconciliation as every competitor. If our arithmetic in this guide is wrong, tell us and we will publish a correction.

OUTLOOK FOR THE REST OF 2026

Telemetry outgrows infrastructure

More traces, profiles, AI-agent events, and high-cardinality context raise both value and volume.

Pricing gets more granular

Elastic already publishes AI token and workflow-execution add-ons; Grafana publishes token and generation meters. “AI included” will increasingly need a usage definition. [7][9]

Control planes matter as much as backends

OpenTelemetry Collectors, routing, sampling, filtering, and ownership metadata become financial infrastructure.

Cold and object storage expand

The practical design is not “keep or drop” but “hot, searchable later, or discard.”

FinOps and SRE converge

The FinOps Foundation’s 2025 report says workload optimization and waste reduction remained the top current priority among respondents. [15]

What this guide is

A vendor-authored benchmark of public observability pricing and a control program, built from desk research. It is not original market research: no survey fieldwork, no customer billing corpus, no interviews.

What we did

On 17 September 2026 we reviewed live vendor pricing pages and documentation for the nine platforms in Section 2, plus published industry surveys. Every price cited comes from a public page listed in Sources.

What we did not do

No contracts signed, no full proofs of concept, no realized enterprise prices. Public list price is not realized enterprise price, and vendors meter compressed or uncompressed volume at different points.

Survey limitations

Grafana's respondents were recruited through its community and events, which may overrepresent open-source users. Sawmills sells telemetry optimization. We cite both because their numbers are visible.

Calculation limits

Decimal units and simple multiplication, excluding tax, and not vendor quotes. Re-run them with billing exports and a controlled proof of concept before making a purchasing decision.

Corrections

Pricing changes frequently. Errors and pricing corrections will be noted in the version history. Version history: first edition, September 2026.

SOURCES

All URLs accessed 17 September 2026

[1] Grafana Labs, Observability Survey 2025 (1,255 respondents)
https://grafana.com/observability-survey/2025/

[2] Sawmills, 2025 State of Observability & Telemetry Report
https://www.sawmills.ai/observability-report-2025

[3] Datadog pricing
https://www.datadoghq.com/pricing/

[4] Datadog Log Management billing
https://docs.datadoghq.com/account_management/billing/log_management/

[5] Splunk Observability pricing
https://www.splunk.com/en_us/products/pricing/observability.html

[6] New Relic pricing and billing
https://newrelic.com/pricing
https://docs.newrelic.com/docs/accounts/accounts-billing/new-relic-one-pricing-billing/new-relic-one-pricing-billing/

[7] Grafana Cloud pricing
https://grafana.com/pricing/

[8] Honeycomb pricing and usage mechanics
https://www.honeycomb.io/pricing
https://docs.honeycomb.io/get-started/manage-costs/how-honeycomb-calculates-usage

SOURCES

All URLs accessed 17 September 2026

[9] Elastic Observability Serverless pricing
https://www.elastic.co/pricing/serverless-observability

[10] Amazon CloudWatch pricing
https://aws.amazon.com/cloudwatch/pricing/

[11] Coralogix pricing
https://coralogix.com/pricing/

[12] KloudMate pricing
https://kloudmate.com/pricing

[13] KloudMate pricing calculator
https://calculator.kloudmate.com/

[14] OpenTelemetry sampling guidance
https://opentelemetry.io/docs/concepts/sampling/

[15] FinOps Foundation, State of FinOps 2025
https://data.finops.org/2025-report/

All prices are public list prices in USD unless stated otherwise. Contract discounts, regional rates, taxes, support tiers, retention, and overages can change the result. Prices were checked against vendor pages on the research date and will drift; re-verify before a purchasing decision.

KloudMate

AI-powered Observability & SRE Copilot

KloudMate replaces and unifies your monitoring and ticketing stack: it cuts alert noise, pinpoints root causes automatically, and hands your engineers everything they need to act. 24×7 operations, without 24×7 headcount.

Our public model: logs and traces by GB, metrics by samples, with no host or user fee. Put it in the same benchmark as everyone else.

SUGGESTED CITATION

KloudMate (2026), Observability Cost Benchmark and Control Guide 2026, first edition, September 2026.

- kloudmate.com