



# FinOps for Streaming and Real-Time Data Platforms

## Optimizing for Value

A summary of the FinOps Foundation paper

Read the full paper at [finops.org](https://finops.org) →

# Real-time platforms power the capabilities the business is paying for

**Fraud prevention. Personalization. Live operational intelligence.**

Real-time data platforms like Kafka, Flink, ClickHouse unlock capabilities that batch architectures cannot deliver.

*They also have a cost profile unlike anything else in the technology estate.*

**Always-on Baseline**

**100%**

of the time, brokers, processors, and serving layers must stay available even when user traffic is quiet.

# Why streaming stresses traditional FinOps

Standard rightsizing heuristics and tag-based allocation models that were built for stateless services and batch workloads fall short.



## Always-on baselines

Brokers and stateful jobs can't be scaled down between jobs. Classic 70–80% utilization targets become unsafe.



## Replication multipliers

Each message produces several disk writes and network transfers across availability zones and regions.



## Stateful operations

Local state, checkpoints, and standby replicas translate directly into storage, I/O, and CPU that quietly compounds.

# FROM SAVINGS TO VALUE



## THE WRONG QUESTION

*“How do we make streaming cheaper?”*



## THE RIGHT QUESTION

*“Is the value-to-cost ratio improving?”*

# The FinOps Domains

## Applied to Streaming & Real-Time Data





DOMAIN

# Understand Usage & Cost

*From one cluster invoice to fair, defensible allocation across topics, consumer groups, and teams.*

Data Ingestion

Allocation

Reporting & Analytics

Anomaly Management

## KEY TAKEAWAYS

**Throughput drives broker and network share.**

**Log size drives storage share (adjusted for replication factor).**

**Metadata ties topics and consumer groups to owning teams.**

## THE PRACTITIONER UNLOCK

FOCUS-conformed billing data makes the model portable across providers and easier for Finance Personas to consume without translation.

*Build three dashboard views: (i)Leadership, (ii)Engineering, (iii)Product - with each answering one question.*



DOMAIN

# Quantify Business Value

*Pair cost with value so the conversation moves from “what does streaming cost” to “is the value-to-cost ratio improving.”*

Unit Economics

Forecasting

Budgeting

Planning & Estimating

KPIs & Benchmarking

## KEY TAKEAWAYS

### Baseline platform cost:

predictable, forecast at high confidence.

### Variable workload cost:

forecast against business volume drivers.

### Cross-region & DR cost:

scales with policy, not traffic.

## PAIR COST → WITH BUSINESS KPIs

Fraud pipeline → **loss prevented per 1,000 decisions**

Personalization → **incremental conversion lift**

Observability → **mean time to detection**



DOMAIN

# Optimize Usage & Cost

*Pull the levers that improve efficiency — without compromising resilience or user experience.*

## Usage Optimization

Rightsize within resilience constraints. Standardize retention by data class. Introduce tiered storage. Clean up unused topics.

**Watch for:** 70–80% utilization targets are unsafe; headroom is partly resilience buffer.

## Architecting & Workload Placement

Co-locate consumers with brokers. Apply compression on the wire. Be selective about cross-region replication.

**Watch for:** cross-region replication is its own cost category; budget it separately.

## Rate Optimization

Cover 60–80% of baseline with commitments. Use on-demand for the rest. Extend to SaaS streaming offerings.

**Pair with:** strong Usage Optimization first; commit only to baselines you actually need.





DOMAIN

# Manage the FinOps Practice

*Keep the work continuous through governance, automation, education, and architecture-level decisions.*

Governance, Policy & Risk

Automation, Tools & Services

Education & Enablement

Exec Strategy Alignment

## GOVERNANCE

Quotas, retention guardrails, and policy-as-code that prevent uncontrolled growth without gatekeeping every change.

## AUTOMATION

Topic lifecycle automation scheduled checks for unused topics, old consumer groups, and orphaned storage.

## EDUCATION

Put cost next to throughput and lag on the same Grafana page. It changes cost from a quarterly conversation to a daily one.

## EXECUTIVE STRATEGY ALIGNMENT

Build-versus-buy. Self-managed vs. SaaS. The architecture decisions that shape what the platform costs in the first place.

## THE CORE ARGUMENT

Streaming usually costs more than batch.  
That's the point.

---

It almost always should because that's what unlocks the capabilities the business is paying for.

**The FinOps question isn't whether streaming is cheap. It's whether the incremental value justifies the spend, and whether the practice is disciplined enough to keep the value-to-cost ratio improving.**

# Unit economics makes the value conversation possible

From “what does streaming cost” to comparable, decision-grade unit costs that Engineering, Product, and Finance can act on together.



## \$/M

### Messages Processed

The most general benchmark; easy to compare across designs and vendors.



## \$/GB

### Retained Per Day

Exposes the impact of retention policy and replication factor.



## \$/Decision

### Served

Ties streaming cost directly to a business-facing outcome.

These metrics let **Engineering**, **Product**, and **Finance** Personas have the same conversation about whether a new feature, consumer, or region is worth what it costs.

## THIS QUARTER

# Three things to start

Leadership-level moves that compound. Re-org or a new tool not required.

### 01

#### **Fund the allocation model**

Make trustworthy allocation a one-quarter deliverable. Without it, every other conversation about streaming cost stalls.

### 02

#### **Set a value-to-cost KPI**

Pick one unit cost metric per streaming pipeline and pair it with a business KPI. Review it at every quarterly business review.

### 03

#### **Split baseline vs. variable**

Forecast and budget the always-on baseline as its own line. A 10% overrun on baseline is a different signal than a 10% overrun on variable spend.

**MONDAY MORNING**

# Three things to take to your team

**01**

## Build a multi-dimensional allocation model

Distribute broker and storage cost by throughput and log size. Tie topics and consumer groups to owning teams through metadata.

**02**

## Put cost next to operational metrics

Cost belongs on the same Grafana page as throughput, lag, and storage. It changes cost from a quarterly conversation to a daily one.

**03**

## Run a quarterly retention review

Lightweight and high-impact. A significant portion of streaming storage cost comes from default or conservative retention settings.

Read the full paper at [finops.org](https://finops.org) →