Redpanda

6 streaming KPIs that make or break AI pipelines

A quick guide to the numbers that keep real-time models fast, accurate, and cost efficient.

Why these metrics matter

Batch dashboards only flag trouble after the fact. Track these streaming KPIs instead, and you'll know—instantly—if your feature pipes are healthy enough for production AI.

Building agents or automated decision-making? Think in events, not bulk streams. Event-driven architecture is how dynamic data reaches models safely and on time.

The KPI table you actually need

Metric	Why it matters for AI/ML	How to track Redpanda	Redpanda vs. Apache Kafka®
Throughput (MB/s, msgs/s)	Keeps feature stores fed during traffic spikes	rpk cluster quotas describe[1]	Thread-per-core design sustains up to 3× Kafka throughput on equal hardware[2]
P99 latency (ms)	Drives inference speed; slow pipes mean stale predictions	Built-in latency dashboard in Redpanda Console	Up to 10× lower P99 vs. JVM-based Kafka[2]
Message durability (ack, ISR)	Lost events create blind spots and retraining cost	Quorum replication + Tiered Storage = zero data loss	Redpanda's Raft avoids slow followers or lost state for superior data safety
End-to-end lag (producer→consumer)	Measures real-world delay from event to action	Rpk group describe lag column[3]	Auto-balancing keeps lag flat as partitions grow
Partition balance (% skew)	Hot partitions throttle model inputs	Continuous Data Balancing enabled by default	Built-in partition balancing removes the need for Cruise Control, cutting a third-party dependency and freeing broker resources[4]
Error rate (fails/sec)	Silent drops distort training data	Dead-letter queues + filter-by-error in Console	Idempotent producers detect and reject duplicate messages, preventing wasted retries and preserving order[5]
Cost per TB per year	Determines ROI and footprint	Redpanda Cloud TCO tools	Up to 6x lower infra cost on average[6]

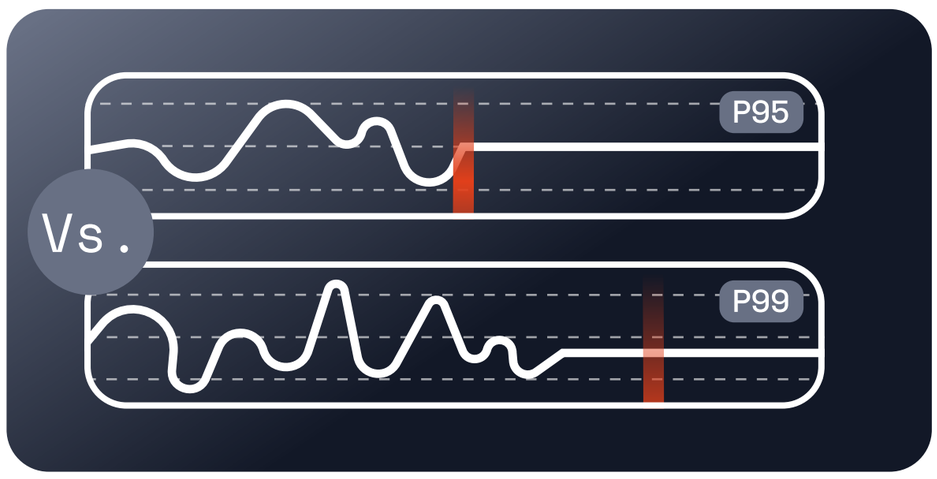
Hit the targets with Redpanda

Keep latency low, lag flat, and data safe
with one binary and built-in tools.

Make these 3 adjustments

1

Set an alert
when lag exceeds
1.5 × steady state.



2

Use the ListKafkaConnections
API endpoint to identify
and troubleshoot
problematic clients.

3

Use Tiered
Storage for long-tail
training features.

Run it now in Redpanda Serverless

SERVERLESS QUICKSTART

TRY THE ENTERPRISE EDITION

Redpanda

Sources:

[1] Redpanda. Manage throughput documentation. Accessed August 7, 2025.

[2] Redpanda. Redpanda vs. Kafka with KRaft: Performance update. May 11 2023. Accessed August 7, 2025.

[3] Redpanda. Rpk group describe documentation. Accessed August 7, 2025.

[4] Redpanda. High Availability documentation. Accessed August 7, 2025.

[5] Redpanda. Idempotent producers documentation. Accessed August 7, 2025.

[6] Redpanda. Redpanda vs. Apache Kafka (TCO Analysis). Published October 18, 2022. Accessed August 7, 2025.