

FINOPS • TEMPLATE

Rightsizing AWS Without Breaking Production: A Data-Driven *Playbook*

KAANSYSTEMS.COM/LIBRARY/RIGHTSIZING-AWS-WITHOUT-BREAKING-
PRODUCTIONSEPTEMBER 16,
2026

ABOUT THIS TEMPLATE

Rightsizing matches AWS resource size to real utilization. Measure first, change one workload at a time behind a tested rollback, and never commit to waste.

THE TEMPLATE

Run through it in order. Each box has a clear "done" state; do not start the compute cuts before the measurement boxes are checked.

Before you touch anything

- ☐ CloudWatch agent installed wherever memory is the binding constraint
- ☐ Metrics collected over a full business cycle (month-end / peak included)
- ☐ Compute Optimizer enabled; recommendations pulled for EC2, ASG, EBS, RDS, Lambda
- ☐ Every candidate resource attributed to a named workload and owner
- ☐ Headroom policy written down (p95 CPU/mem under ~70%; more for latency-SLO services)
- ☐ T-family burstables checked for CPU-credit balance, not just average CPU

Making the change (one workload at a time)

- ☐ Change lives in the launch template / ASG / IaC, not a console click
- ☐ Previous version or snapshot confirmed as a working rollback
- ☐ Applied to one workload, in a low-traffic window for anything stateful
- ☐ p95/p99 latency and error rate watched before and after, for a full cycle
- ☐ Result backported to IaC so drift detection stays quiet

Storage and database sweep

- ☐ All gp2 volumes migrated to gp3
- ☐ Unattached EBS volumes snapshotted and deleted
- ☐ Orphaned snapshots and unreferenced AMIs retired on a retention policy
- ☐ S3 lifecycle or Intelligent-Tiering applied to cold data
- ☐ RDS rightsized last, treated as a production change (window + rollback)

Only then

- ☐ Steady-state baseline re-measured after the cuts
- ☐ Savings Plans / RIs sized against the new, smaller baseline — never the old one

— HOW TO USE IT

Rightsizing is not a one-time event. Workloads grow back into their instances, someone launches a fresh gp2 volume, a new service ships over-provisioned "to be safe." Re-run the measurement pass on the same cadence as the [30-day cost cleanup](#) — a couple of times a year — and the estate stays close to its true workload cost instead of drifting back up.

Keep the discipline even when the numbers are small. The habit that saves you is not the single dramatic cut; it is that every change goes through measurement, a rollback, and a watch window, so the day you rightsize something load-bearing, the muscle is already there. The teams that break production rightsizing are the ones who did it carefully ten times, got comfortable, and skipped the rollback on the eleventh.