



Vantalect

AWS COST · SAMPLE DELIVERABLE

Quick Wins

Week 3 of 6: Quick wins · prepared for Tidepost, Inc.

Prepared by Charles Sieg · charles.sieg@vantalect.com · vantalect.com

April 24, 2026 · AWS Cost Analysis & Optimization: The Optimization (six weeks)

SAMPLE DELIVERABLE

This is a sample deliverable. Tidepost, Inc. is a fictional company: its people, products, users, accounts, investors, and figures are invented for this sample, and any resemblance to a real company, person, product, or publication is coincidental. AWS prices are list prices from the AWS Price List API retrieved on October 13, 2026, and from the public AWS Price List bulk API on October 18, 2026.

Contents

Summary	2
What went in	4
Change by change	6
The evidence the auditor can read	9
The changes, one at a time	10
SM-2: Notebooks stop themselves	10
EC2-2: Developer instances sleep at night	11
NET-1: Gateway endpoints for S3 and DynamoDB	11
NET-2: Fewer public IPv4 addresses	12
EC2-3: The API tier's minimum, with target tracking	12
LOG-1: INFO in production, not DEBUG	13
LOG-3: Custom metrics without the per-post dimension	13
EBS-1: gp2 volumes to gp3	13
S3-4: Old exports and CI artifacts expire	14
EBS-2: Unattached volumes archived, then deleted	14
ECR-1: An ECR lifecycle policy	15
TR-1: Transcribe only videos with speech	15
KIN-1: Kinesis retention cut back	15
What cannot be undone	16
The Well-Architected questions these changes close	16
Decided at the readout, and not made	17
Next week	18

Summary

Dev Raman and Kofi Brandvold: in week 3 of the Optimization, from Monday, April 20, to Friday, April 24, Tidepost's engineers put **thirteen changes in place** across three accounts (tp-dev, tp-prod, and tp-sandbox-m1). They are estimated to save **\$228,238 a year**, 10.8% of the April bill annualized (\$2,110,762). None of it shows on an invoice yet: the first change went in on Tuesday, April 21, and the Cost and Usage Report needs seven full days after a change before it can be read.

Every dollar here is an estimate. It is the Analysis’s price for the change at the April 6 usage, annualized, at list price. The week-six measurement report, on May 15, measures each change on the Cost and Usage Report, and the figure it prints will differ from the one here for a reason it will name. Of the 40 changes the Analysis priced, thirteen are in place at this report’s date. The largest change of week 4 is SM-1a (The embedding endpoint on one GPU), at \$97,674 a year ([Next week](#)).

Every change was approved by the owner of the workload at the April 17 readout, made by a Tidepost engineer through a pull request and Tidepost’s pipeline, and reviewed by Vantalect, which holds a read-only role and writes nothing to production. Every change has a way back, written beside it. Three cannot be fully reversed, because they remove something that does not come back, and they are named in the section [What cannot be undone](#).

Week 3 in numbers. Savings are annualized estimates at list price; the measurement is week six.

Week 3 at a glance	Result
Changes in place, Monday, April 20, to Friday, April 24	13
Estimated savings, a year	\$228,238
The same, a month	\$19,020
Share of the April bill annualized	10.8%
Accounts touched	3
Owners who approved	5
Well-Architected questions with a change in place	3 of 5
Well-Architected questions closed	0 of 5
Measured savings	none until week six

What the sponsor needs to do

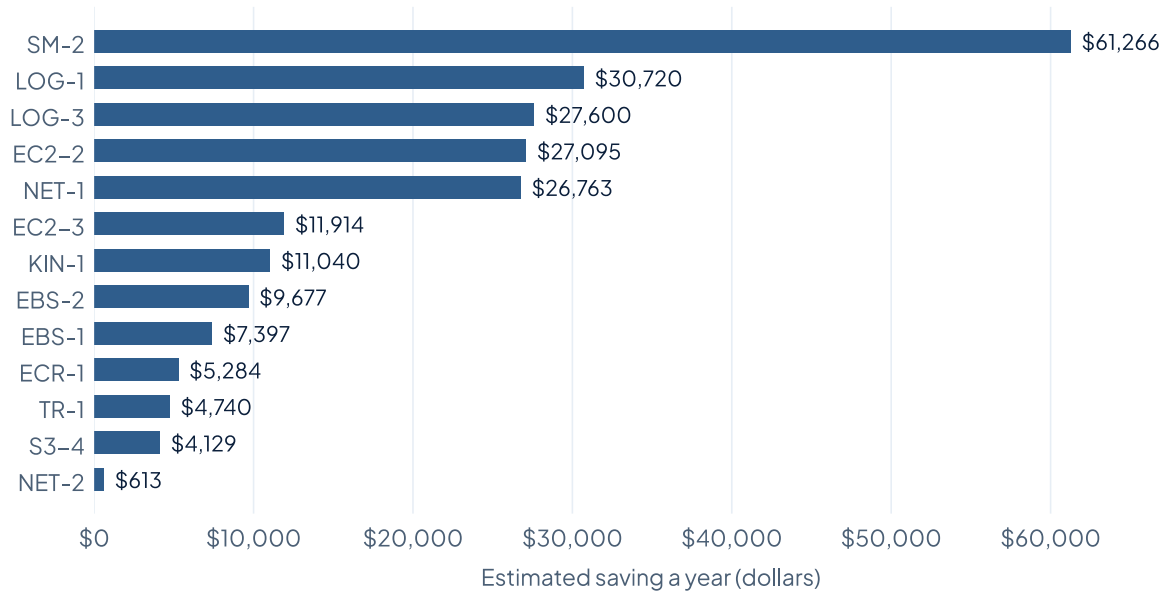
Dev Raman: nothing is needed to keep this week’s savings. Week 4 starts Monday, April 27, with the embedding endpoint (Yuki Sorensen’s team) and the production nodes (Kofi Brandvold’s), and includes the Compute Savings Plan on April 30, which Gwen Tallis owns and approves in writing before it is bought. Please ask Yuki Sorensen’s team to watch the endpoint swap on April 27; the swap interrupts nothing.

What went in

Seven kinds of change went in this week. The largest by value is waste, \$92,009 a year across six changes. The four largest changes by size are SM-2, LOG-1, LOG-3, EC2-2, and together they are \$146,681 of the week's \$228,238 ([the figure below](#)).

The week's changes by kind. The estimates add up to the week total to the dollar.

Kind of change	IDs	Changes	Estimate, a year	Share of week 3
Waste	SM-2, NET-2, S3-4, EBS-2, ECR-1, KIN-1	6	\$92,009	40.3%
Logging	LOG-1, LOG-3	2	\$58,320	25.6%
Schedules	EC2-2	1	\$27,095	11.9%
Networking	NET-1	1	\$26,763	11.7%
Rightsizing	EC2-3	1	\$11,914	5.2%
Storage	EBS-1	1	\$7,397	3.2%
Moderation	TR-1	1	\$4,740	2.1%
Total		13	\$228,238	100.0%



Estimated saving a year for each change in place in week 3.

When each change went in. Tuesday, April 21, was the first day of the week's changes and the first day any of them could start being read on the Cost and Usage Report.

Day	In place	Changes	Estimate, a year
Tuesday, April 21	SM-2, EC2-2, NET-1, NET-2	4	\$115,737
Wednesday, April 22	EC2-3, LOG-1, LOG-3, EBS-1	4	\$77,631
Thursday, April 23	S3-4, EBS-2, ECR-1	3	\$19,090
Friday, April 24	TR-1, KIN-1	2	\$15,780
Total		13	\$228,238

Change by change

Each row names the change, the account, the owner, and the date, then what the resource looked like before, what it looks like now, the estimate, and the way back. "Before" is the line of the April 6 bill that the change touches.

Every change in place in week 3, in the order it went in. Estimates are annualized at list price; nothing is measured.

Change	Before	After	Way back	Estimate, a year
SM-2 Notebooks stop themselves. tp-sandbox-ml; Yuki Sorensen; Tuesday, April 21.	Notebooks, never stopped: 1 ml.g5.12xlarge and 2 ml.g5.2xlarge, always on, at \$7.09 and \$1.52 an hour.	Notebook auto-stop after 60 idle minutes	Detach the lifecycle configuration and restart the notebooks. Nothing is deleted, and a stopped notebook starts again from its saved volume.	\$61,266
EC2-2 Developer instances sleep at night. tp-dev; Kofi Brandvold; Tuesday, April 21.	Developer instances: 12 m5.2xlarge, On-Demand, always on, at \$0.384 an hour.	Developer instances stop outside working hours	Remove the schedule in one pull request. Instances start at the next run, and their volumes keep their data throughout.	\$27,095
NET-1 Gateway endpoints for S3 and DynamoDB. tp-prod; Kofi Brandvold; Tuesday, April 21.	NAT gateways: 3 gateways, 60.5 TB processed, at \$0.045 an hour and \$0.045 a GB.	S3 and DynamoDB gateway endpoints	Delete the endpoints. Their route-table entries go with them, and traffic returns to the NAT gateways at once.	\$26,763
NET-2 Fewer public IPv4 addresses. tp-prod; Kofi Brandvold; Tuesday, April 21.	Public IPv4 addresses: 20 addresses, at \$0.005 an hour.	Public IPv4 addresses from 20 to 6	Allocate a new address. A released address cannot be taken back, which is why each one was checked against the allow lists that partners hold before it was released.	\$613
EC2-3 The API tier's minimum, with target tracking. tp-prod; Marcus Oyelaran; Wednesday, April 22.	API tier: 9 c6i.4xlarge, On-Demand, always on, at \$0.680 an hour.	API tier minimum from 9 to 6 (two a zone), with target tracking	Raise the minimum in tidepost-cdk. The group launches the missing instances within minutes.	\$11,914

Change	Before	After	Way back	Estimate, a year
LOG-1 INFO in production, not DEBUG. tp-prod; Marcus Oyelaran; Wednesday, April 22.	Log ingestion (the API logs at DEBUG): 6.6 TB a month, at \$0.50 a GB.	API log level from DEBUG to INFO in production	Set the level back to DEBUG in the service configuration (one pull request and a deploy), or switch it on for one service during an incident.	\$30,720
LOG-3 Custom metrics without the per-post dimension. tp-prod; Marcus Oyelaran; Wednesday, April 22.	Custom metrics (a per-post dimension): 12,000 metrics, at the published tiers.	Custom metrics from 12,000 to 3,000 (the per-post dimension dropped)	Add the dimension back in the publisher and deploy. New series start from that deploy; the dropped dimension is not rebuilt for earlier hours.	\$27,600
EBS-1 gp2 volumes to gp3. tp-prod; Kofi Brandvold; Wednesday, April 22.	gp2 volumes, attached: 30.1 TB, at \$0.10 a GB-month.	gp2 to gp3 (30.1 TB)	Modify the volume back to gp2 once AWS's cool-down between modifications of one volume has passed. The size does not change.	\$7,397
S3-4 Old exports and CI artifacts expire. tp-prod; Kofi Brandvold; Thursday, April 23.	Old exports and CI artifacts sit in S3 Standard with no expiration rule, inside the S3 Standard line of the bill.	Old exports and CI artifacts expired (16 TB)	No way back for an object once it has expired. If a prefix turns out to be needed, the rule is removed before the next daily lifecycle run, which is why the list was signed first.	\$4,129
EBS-2 Unattached volumes archived, then deleted. tp-prod; Kofi Brandvold; Thursday, April 23.	gp2 volumes, unattached: 9 TB, at \$0.10 a GB-month.	Unattached volumes archived as snapshots, then deleted (9 TB)	Restore a volume from the archived snapshot. A restore from the archive tier takes a day or more, which is the trade for its price.	\$9,677
ECR-1 An ECR lifecycle policy. tp-prod; Kofi Brandvold; Thursday, April 23.	ECR image storage: 5 TB, at \$0.10 a GB-month.	ECR lifecycle policy (5 TB to 0.7 TB)	Remove the policy to stop further expiry. An expired image is rebuilt by the pipeline from source.	\$5,284
TR-1 Transcribe only videos with speech. tp-prod; Amara Delacroix; Friday, April 24.	Transcribe, every video: 173,250 minutes a month, at \$0.0001 a second.	Transcribe only videos with detected speech	Turn the check off with its feature flag. Every video is transcribed again from the next upload.	\$4,740

Change	Before	After	Way back	Estimate, a year
KIN-1 Kinesis retention cut back. tp-prod; Lior Haddad; Friday, April 24.	Long-term retention (365 days; data older than seven days): 40,000 GB stored, at \$0.023 a GB-month.	Kinesis retention from 365 days to 7 (the raw stream is already in S3)	IncreaseStreamRetentionPeriod raises the period again, for new records only. Records that have aged out of the stream do not come back; they remain in S3.	\$11,040
Total		13 changes		\$228,238

The evidence the auditor can read

Tidepost’s SOC 2 Type 2 observation window opened April 1 and runs to September 30, so every change in this report was made inside it. The change-management control (CC8.1) has been operating since March 27: Production changes through a pull request and the pipeline: read-only permission sets in tp-prod for developers, machine learning, and data, console-built resources imported into the CDK, and drift reports on console changes. The auditor tests a sample of the changes in the window; Vantalect does not select the sample and gives no opinion on what the auditor finds. What follows is where each record is, so that Sasha Ilyin can produce it.

The five records that evidence one change. The auditor tests the control; this table is not a test of it.

Record	Where it lives	What it shows
The pull request	The repository named in the next table	The diff, the reviewer, the approval, and the change’s identifier in its title
The pipeline run	Tidepost’s pipeline, deploying through a role per repository	The build, the deploy, who merged, and the time
The CloudFormation drift report	The stack’s drift detection after the deploy	That the account matches the code once the change is in
The API call	The organization trail in tp-log-archive	The call that made the change, and the pipeline role that made it
The change window	The date and window class in the next table	When the change was made and whether anyone was told

The repository, date, and window class for every change in week 3.

Change	Repository of the pull request	Date	Window	Approved by
SM-2	tidepost-cdk	Tuesday, April 21	Standard	Yuki Sorensen, at the April 17 readout
EC2-2	tidepost-cdk	Tuesday, April 21	Standard	Kofi Brandvold, at the April 17 readout
NET-1	tidepost-cdk	Tuesday, April 21	Standard	Kofi Brandvold, at the April 17 readout
NET-2	tidepost-cdk	Tuesday, April 21	Standard	Kofi Brandvold, at the April 17 readout
EC2-3	tidepost-cdk	Wednesday, April 22	Standard	Marcus Oyelaran, at the April 17 readout

Change	Repository of the pull request	Date	Window	Approved by
LOG-1	the API's configuration in tidepost-cdk	Wednesday, April 22	Standard	Marcus Oyelaran, at the April 17 readout
LOG-3	the API's repository	Wednesday, April 22	Standard	Marcus Oyelaran, at the April 17 readout
EBS-1	tidepost-cdk	Wednesday, April 22	Standard	Kofi Brandvold, at the April 17 readout
S3-4	tidepost-cdk	Thursday, April 23	Standard	Kofi Brandvold, at the April 17 readout
EBS-2	tidepost-cdk	Thursday, April 23	Standard	Kofi Brandvold, at the April 17 readout
ECR-1	tidepost-cdk	Thursday, April 23	Standard	Kofi Brandvold, at the April 17 readout
TR-1	the moderation service's repository	Friday, April 24	Standard	Amara Delacroix, at the April 17 readout
KIN-1	tidepost-cdk	Friday, April 24	Standard	Lior Haddad, at the April 17 readout

Standard. Nothing users see is interrupted, so the day the pipeline made the change is its window.

The changes, one at a time

SM-2: Notebooks stop themselves

Owner: Yuki Sorensen. **Account:** tp-sandbox-ml. **In place:** Tuesday, April 21. **Estimate:** \$61,266 a year (\$5,106 a month). **What changed:** Notebook auto-stop after 60 idle minutes.

How it was made. A lifecycle configuration with an idle check is attached to each notebook instance in tp-sandbox-ml. A notebook with no kernel activity for the idle period stops itself; a stopped notebook keeps its volume and its files. **The estimate** is 1 ml.g5.12xlarge at \$7.09 and 2 ml.g5.2xlarge at \$1.52, from 730 hours to 220 and 240. **Why it was safe:** Kernel activity: the notebooks were idle 74% of their hours.

Way back. Detach the lifecycle configuration and restart the notebooks. Nothing is deleted, and a stopped notebook starts again from its saved volume. **What the owner watches.** Yuki Sorensen's team starts a stopped notebook when it needs one. Week five makes the same

default apply to every new notebook (GR-3). **Well-Architected.** This change is on the list for COST06 (closes April 27, when SM-1a lands) and SUS02 (closes April 27, when SM-1a lands), accepted in writing and taken up by the cost engagement.

Next.

- **SM-1a** (planned for Monday, April 27): the endpoint side of the SageMaker work; together they close the two Well-Architected questions that wait on them.
- **GR-3** (built in week 5): makes auto-stop the default for every new notebook.

EC2-2: Developer instances sleep at night

Owner: Kofi Brandvold. **Account:** tp-dev. **In place:** Tuesday, April 21. **Estimate:** \$27,095 a year (\$2,258 a month). **What changed:** Developer instances stop outside working hours.

How it was made. A schedule in the stack stops every instance in tp-dev outside working hours and starts it before the workday. An exemption tag keeps one instance running for an engineer who asks, and the tag is visible in the cost allocation data. **The estimate** is 12 m5.2xlarge at \$0.384, from 730 hours to 240. **Why it was safe:** CPU by hour: under 2% outside 8 AM to 6 PM on 11 of 12.

Way back. Remove the schedule in one pull request. Instances start at the next run, and their volumes keep their data throughout. **What the owner watches.** Engineers tell the platform team when a machine is still asleep at the start of the day; the schedule's start time is a parameter, not code. **Well-Architected.** This change is on the list for COST06 (closes April 27, when SM-1a lands), accepted in writing and taken up by the cost engagement.

Next.

- **EC2-1a** (deferred to Monday, June 1): the old staging is rebuilt smaller in tp-staging.
- **EC2-1b** (deferred to Monday, June 8): staging then follows the same weekday schedule.

NET-1: Gateway endpoints for S3 and DynamoDB

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Tuesday, April 21. **Estimate:** \$26,763 a year (\$2,230 a month). **What changed:** S3 and DynamoDB gateway endpoints.

How it was made. Gateway endpoints for S3 and DynamoDB are added to the production VPC and associated with the private route tables, so same-Region traffic to those two services stops passing through the NAT gateways. Gateway endpoints carry no hourly or data charge. The organization's S3 data perimeter (a resource control policy from the Foundation) still applies to every request. **The estimate** is 80% of 60.5 TB processed at \$0.045 a GB. **Why it was safe:** VPC flow logs: 80% of NAT bytes go to S3 and DynamoDB prefixes.

Way back. Delete the endpoints. Their route-table entries go with them, and traffic returns to the NAT gateways at once. **What the owner watches.** A gateway endpoint carries only traffic to the same Region, so traffic to a bucket in another Region stays on the NAT gateways. The estimate prices what the flow logs show going to same-Region prefixes; week six reads what actually moved.

NET-2: Fewer public IPv4 addresses

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Tuesday, April 21. **Estimate:** \$613 a year (\$51 a month). **What changed:** Public IPv4 addresses from 20 to 6.

How it was made. The Elastic IP addresses attached to stopped instances and to the old NAT gateway are released. The addresses that remain belong to the load balancers and the live NAT gateways. **The estimate** is 14 addresses $\times$ \$0.005 $\times$ 730 hours. **Why it was safe:** Elastic IPs on stopped instances and the old NAT.

Way back. Allocate a new address. A released address cannot be taken back, which is why each one was checked against the allow lists that partners hold before it was released. **What the owner watches.** A partner that allow-lists one of Tidepost's addresses would see failures at once; none of the released addresses was on a list.

EC2-3: The API tier's minimum, with target tracking

Owner: Marcus Oyelaran. **Account:** tp-prod. **In place:** Wednesday, April 22. **Estimate:** \$11,914 a year (\$993 a month). **What changed:** API tier minimum from 9 to 6 (two a zone), with target tracking.

How it was made. The API tier's Auto Scaling group takes a smaller minimum, two instances in each of the three zones, and a target-tracking policy on average CPU. It adds instances for the daily peak and sheds them at night, one zone at a time. **The estimate** is From 9 to 7 c6i.4xlarge on average at \$0.680. **Why it was safe:** Request rate and CPU by hour: peaks need 8 instances, nights need 4.

Way back. Raise the minimum in tidepost-cdk. The group launches the missing instances within minutes. **What the owner watches.** Marcus Oyelaran's team watches request latency and CPU against the service-level alarms in place since March.

Next.

- **EC2-4** (planned for Friday, May 1): moves the same tier to Graviton, priced after this change.

LOG-1: INFO in production, not DEBUG

Owner: Marcus Oyelaran. **Account:** tp-prod. **In place:** Wednesday, April 22. **Estimate:** \$30,720 a year (\$2,560 a month). **What changed:** API log level from DEBUG to INFO in production.

How it was made. The API's production log level is set to INFO in the service configuration. DEBUG stays available for one service at a time, switched on by a pull request for the length of an investigation. **The estimate** is From 6.6 TB to 1.6 TB a month at \$0.50 a GB. **Why it was safe:** Log groups by level: 76% of bytes are DEBUG.

Way back. Set the level back to DEBUG in the service configuration (one pull request and a deploy), or switch it on for one service during an incident. **What the owner watches.** The log groups by level show whether any INFO line is missing a field an on-call engineer relies on; the on-call rotation reviews that in the next incident review.

Next.

- **LOG-2** (planned for Tuesday, May 5): shortens how long the same log groups are kept.
- **GR-1** (built in week 5): sets the level and retention defaults on every new log group.

LOG-3: Custom metrics without the per-post dimension

Owner: Marcus Oyelaran. **Account:** tp-prod. **In place:** Wednesday, April 22. **Estimate:** \$27,600 a year (\$2,300 a month). **What changed:** Custom metrics from 12,000 to 3,000 (the per-post dimension dropped).

How it was made. The metric publisher stops emitting the post identifier as a dimension. The route, status, and queue aggregates remain, and no dashboard or alarm reads any of the series that stopped. **The estimate** is The published metric tiers before and after. **Why it was safe:** 9,000 metrics carry a post id; no dashboard or alarm reads them.

Way back. Add the dimension back in the publisher and deploy. New series start from that deploy; the dropped dimension is not rebuilt for earlier hours. **What the owner watches.** Marcus Oyelaran's team checks the dashboards and alarms for missing data after the deploy.

Next.

- **GR-5** (in place since March 13, from the Well-Architected Remediation): budgets and anomaly alerts would notice a metric count that climbs again.

EBS-1: gp2 volumes to gp3

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Wednesday, April 22. **Estimate:** \$7,397 a year (\$616 a month). **What changed:** gp2 to gp3 (30.1 TB).

How it was made. Each gp2 volume is changed to gp3 in place with ModifyVolume while it stays attached: no detach, no restart. The same pull request changes the launch templates, so a replaced instance comes back as gp3 and the CloudFormation drift report stays clean. Nothing is provisioned above the gp3 baseline, because no volume needs it. **The estimate** is 30.1 TB × (\$0.10 – \$0.08). **Why it was safe:** CloudWatch over 30 days: no volume peaked above 3,000 IOPS or 125 MiB/s, so gp3's included baseline covers every one with nothing provisioned.

Way back. Modify the volume back to gp2 once AWS's cool-down between modifications of one volume has passed. The size does not change. **What the owner watches.** The volumes' IOPS and throughput stay inside the gp3 baseline, and the CloudWatch read in the evidence is the reason.

S3–4: Old exports and CI artifacts expire

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Thursday, April 23. **Estimate:** \$4,129 a year (\$344 a month). **What changed:** Old exports and CI artifacts expired (16 TB).

How it was made. A lifecycle expiration rule is added on the prefixes of old exports and CI artifacts, from a list Kofi Brandvold signed after reading the last-accessed dates. The rule expires the objects; there is no copy. **The estimate** is 16 TB at \$0.021. **Why it was safe:** Last-accessed dates: none read since 2025.

Way back. No way back for an object once it has expired. If a prefix turns out to be needed, the rule is removed before the next daily lifecycle run, which is why the list was signed first. **What the owner watches.** Nothing was read from these prefixes since the year before, so a missing object would be the first read in more than a year.

Next.

- **S3–3** (planned for Tuesday, May 5): the next storage-class change.
- **S3–2** (deferred to Wednesday, May 20): the largest lifecycle change, and the one that waits on counsel.

EBS–2: Unattached volumes archived, then deleted

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Thursday, April 23. **Estimate:** \$9,677 a year (\$806 a month). **What changed:** Unattached volumes archived as snapshots, then deleted (9 TB).

How it was made. Each unattached volume is snapshotted, the snapshot is moved to the snapshot archive tier, and then the volume is deleted. Only volumes detached for longer than the evidence's period were on the list, and the owner signed it first. **The estimate** is 9 TB from \$0.10 to \$0.0125 (archive tier). **Why it was safe:** Detached for more than 30 days.

Way back. Restore a volume from the archived snapshot. A restore from the archive tier takes a day or more, which is the trade for its price. **What the owner watches.** An engineer who finds a volume missing asks the platform team, and the ledger of this change names the snapshot to restore.

Next.

- **EBS-3** (planned for Monday, May 4): the snapshot side of the same cleanup.

ECR-1: An ECR lifecycle policy

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Thursday, April 23. **Estimate:** \$5,284 a year (\$440 a month). **What changed:** ECR lifecycle policy (5 TB to 0.7 TB).

How it was made. A lifecycle policy on each repository expires images that have not been pulled within the period in the evidence. A second rule keeps the release-tagged images. **The estimate** is 4.3 TB at \$0.10. **Why it was safe:** Images never pulled in 90 days.

Way back. Remove the policy to stop further expiry. An expired image is rebuilt by the pipeline from source. **What the owner watches.** A deploy that needs an expired image fails its pull, and the pipeline rebuilds it; Kofi Brandvold's team sees that in the pipeline run.

TR-1: Transcribe only videos with speech

Owner: Amara Delacroix. **Account:** tp-prod. **In place:** Friday, April 24. **Estimate:** \$4,740 a year (\$395 a month). **What changed:** Transcribe only videos with detected speech.

How it was made. A voice-activity check runs on each upload's audio before the Transcribe job is submitted. A video with no detected speech skips the job, and the check errs toward transcribing when it is unsure. The frame path and the image checks are untouched. **The estimate** is 173,250 minutes at \$0.006 a minute, less the 38% with no speech. **Why it was safe:** A week of Transcribe output: 38% of transcripts were empty.

Way back. Turn the check off with its feature flag. Every video is transcribed again from the next upload. **What the owner watches.** Amara Delacroix's team samples the videos the check skipped each week against a human listen.

KIN-1: Kinesis retention cut back

Owner: Lior Haddad. **Account:** tp-prod. **In place:** Friday, April 24. **Estimate:** \$11,040 a year (\$920 a month). **What changed:** Kinesis retention from 365 days to 7 (the raw stream is already in S3).

How it was made. The raw stream's retention period is reduced with `DecreaseStreamRetentionPeriod`. Data older than the new period is deleted by Kinesis at once, and it already sits in S3: Firehose has landed the raw stream there since launch, which is why no consumer reads past the new period. **The estimate** is 40,000 GB at \$0.023 a GB-month. **Why it was safe:** No consumer reads past seven days (`GetRecords` by iterator age).

Way back. `IncreaseStreamRetentionPeriod` raises the period again, for new records only. Records that have aged out of the stream do not come back; they remain in S3. **What the owner watches.** Lior Haddad's team watches the consumers' iterator age. Week five turns the cap into a guardrail (GR-6). **Well-Architected.** This change is on the list for SUS04 (closes May 6, when S3-1 lands), accepted in writing and taken up by the cost engagement.

Next.

- **KIN-2** (planned for Thursday, April 30): reduces the shard count of the same stream.
- **GR-6** (built in week 5): caps retention on every stream.

What cannot be undone

Three changes cannot be fully reversed

S3-4 expires objects, and an expired object is gone: Kofi Brandvold signed the list of prefixes after the last-accessed dates were read, and the rule can still be removed before its next daily run. **KIN-1** deleted the stream's records older than the new retention at once; they are in S3, and the stream cannot take them back. **NET-2** released public IP addresses, and a released address cannot be re-acquired; each was checked against the allow lists partners hold first. Every other change this week reverses with a pull request or the step beside it.

The Well-Architected questions these changes close

The Well-Architected Review ended on April 3 with five questions accepted in writing, because their fixes are cost changes and a cost change is counted by exactly one engagement. A question closes when the last change on its list is in place. None of the five closes this week: this week's changes are the first on three of the lists.

The five questions accepted in writing, with the dates the facts give. Tidepost saves its answers in the Well-Architected Tool at the May 14 milestone.

Question	Signed by	Changes on its list	In place at April 24	Closes
COST06 How do you meet cost targets when you select resource type, size and number?	Gwen Tallis, March 10	EC2-2, SM-1a, SM-2	EC2-2, SM-2	April 27, when SM-1a lands
SUS02 How do you align cloud resources to your demand?	Yuki Sorensen, March 11	SM-1a, SM-2	SM-2	April 27, when SM-1a lands
COST07 How do you use pricing models to reduce cost?	Gwen Tallis, March 10	SP-1	none	April 30, when SP-1 lands
PERF03 How do you store, manage, and access data in your workload?	Lior Haddad, March 9	S3-1, ATH-1	none	May 6, when S3-1 lands
SUS04 How do you take advantage of data management policies and patterns to support your sustainability goals?	Lior Haddad, March 9	KIN-1, LOG-2, S3-1, S3-3	KIN-1	May 6, when S3-1 lands

Decided at the readout, and not made

Three changes were priced and then declined at the April 17 readout. They are not in this report's estimate and they are not counted. Each decision stands on its owner's reason.

The changes declined at the readout. They stay in the register and are not counted.

Change	Decided by	Reason	Estimate, a year
MOD-2 Frame sampling from one frame a second to one every two seconds	Amara Delacroix	after MOD-1 the frame path alone carries most videos, and halving it halves what it sees	\$37,422
RDS-2 Writer from db.r6g.4xlarge to db.r6g.2xlarge (Multi-AZ)	Dev Raman	growth after the round, and a resize is a failover inside the observation window	\$15,759
OS-1 OpenSearch data nodes from r6g.2xlarge.search to r6g.xlarge.search	Lior Haddad	the search team is midway through a k-NN migration	\$8,778

Next week

Week 4 runs April 27 to May 1. The facts give these dates, each owned by the person named.

The changes planned for week 4.

Change	Planned	Owner	Estimate, a year
EKS-1 Karpenter consolidates the production nodes	Monday, April 27	Kofi Brandvold	\$26,911
SM-1a The embedding endpoint on one GPU	Monday, April 27	Yuki Sorensen	\$97,674
CACHE-1 Smaller ElastiCache nodes	Tuesday, April 28	Kofi Brandvold	\$28,768
EC2-5 Transcoders scale on queue depth	Tuesday, April 28	Kofi Brandvold	\$19,763
MOD-1 Rekognition Video where it earns its cost	Tuesday, April 28	Amara Delacroix	\$83,160
CDN-1 A leaner rendition ladder	Wednesday, April 29	Bea Okafor	\$55,443
DDB-1 Provisioned capacity on the two steady tables	Thursday, April 30	Lior Haddad	\$25,058
KIN-2 Fewer Kinesis shards	Thursday, April 30	Lior Haddad	\$25,795
RDS-1 Smaller read replicas	Thursday, April 30	Dev Raman	\$7,866
SP-1 The Compute Savings Plan	Thursday, April 30	Gwen Tallis	\$23,764
EC2-4 The API tier on Graviton	Friday, May 1	Marcus Oyelaran	\$6,132

Week 5 (May 4 to May 8) is planned to put seven more in place (EBS-3, SM-1b, LOG-2, S3-3, ATH-1, CDN-1b, S3-1), and Gwen Tallis decides SP-2 on May 5. Five more have dates after the engagement's planned changes: CACHE-2 (Redis OSS to Valkey on the same nodes) is deferred to May 19; S3-2 (Originals to S3 Glacier Instant Retrieval) is deferred to May 20; EC2-1a (Staging rebuilt smaller in tp-staging) is deferred to June 1; EC2-1b (Staging on weekdays only) is deferred to June 8; SM-3 (The moderation classifier on Inferentia2) is deferred to July 13. Each is deferred for a reason the facts give, and none is dropped.

AWS, Amazon Web Services, AWS Graviton, AWS Inferentia, and the AWS service names are trademarks of Amazon.com, Inc. or its affiliates. PostgreSQL is a registered trademark of the PostgreSQL Community Association of Canada. Kubernetes is a registered trademark of The Linux Foundation. Redis is a registered trademark of Redis Ltd. Valkey is a trademark of the Linux Foundation. OpenSearch is a registered trademark of Amazon.com, Inc. SOC 2 is a registered trademark of the American Institute of Certified Public Accountants. CIS Benchmarks is a trademark of the Center for Internet Security, Inc. GitHub is a trademark of GitHub, Inc. Jupyter is a trademark of LF Charities. Vantalect is not affiliated with or endorsed by any of them. Tidepost, Inc. is fictional.