

AWS COST · SAMPLE DELIVERABLE

Data Lifecycle, Telemetry, and Guardrails

Week 5 of 6: Architecture and guardrails · prepared for
Tidepost, Inc.

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

May 8, 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	3
What went in	4
Change by change	7
Telemetry: Parquet and Athena	9
Storage tiers: renditions now, originals on May 20	9
Renditions to Glacier Instant Retrieval, priced against Standard-IA and Intelligent-Tiering	9
Originals to Glacier Instant Retrieval: S3-2, deferred to May 20	11
Logs: LOG-2, and the audit logs it leaves alone	12
Snapshots: EBS-3 and AWS Backup	13
The SageMaker Savings Plan: SP-2	13
Guardrails	13
Cost avoidance, not counted	17
The evidence the auditor can read	19
The changes, one at a time	20
SM-1b: The embedding endpoint scales on its backlog	20
EBS-3: Snapshot chains past their keep period deleted	21
LOG-2: Shorter application log retention, counsel's copy first	21
S3-3: Old renditions to Glacier Instant Retrieval	22
S3-1: Telemetry to Parquet at ingest	23
ATH-1: Athena on partitioned Parquet	23
CDN-1b: One rendition fewer to transcode	24
What cannot be undone	24
The Well-Architected questions	24
Still ahead, and how week six measures	25

Summary

Dev Raman and Kofi Brandvold: in week 5 of the Optimization, Monday, May 4, to Friday, May 8, Tidepost’s engineers put **seven changes in place**, estimated to save **\$152,131 a year**, and built **five guardrails** that keep the waste from coming back, beside the two the engagement inherited (GR-5 and GR-7). With weeks 3 and 4, **31 changes are in place**, an estimated **\$780,703 a year**, **37.0% of the April bill** annualized (\$2,110,762). That is every change the facts give an in-place date. Nothing here is measured; the week-six report on May 15 measures each change on the Cost and Usage Report.

The week was about **data that gets old**. Telemetry went to Parquet at ingest (S3-1) and the queries followed (ATH-1); renditions older than 60 days went to Glacier Instant Retrieval (S3-3); application logs moved to a shorter retention after counsel’s copy of January to March was safe (LOG-2); and the snapshot chains the old cron job left behind were deleted, with AWS Backup the only schedule from here (EBS-3). Gwen Tallis also declined the SageMaker Savings Plan (SP-2) on Tuesday, May 5, because she wants the Inferentia2 move first ([SP-2](#)).

Cost avoidance, not counted. If usage grows from today’s monthly users to the 1,500,000 Tidepost plans for December, the old configurations these changes replaced would have grown with it. We estimate that growth at **\$140,000 to \$230,000 a year** at the December volume. It is cost avoidance, it is not counted toward anything, and it is a model with stated assumptions ([Cost avoidance](#)). A private pricing negotiation with CloudFront, which Gwen Tallis owns, is named in the roadmap and is not counted or estimated here.

Week 5 in numbers. Savings are annualized estimates at list price; the measurement is week six. The avoidance range is not savings.

Week 5 at a glance	Result
Changes in place, May 4 to May 8	7
Estimated savings, a year	\$152,131
The same, a month	\$12,678
In place so far (weeks 3 to 5)	31
Estimated savings so far, a year	\$780,703
Share of the April bill annualized	37.0%
Guardrails built in week 5, one owner each	5
Guardrails inherited from earlier engagements	2

Week 5 at a glance	Result
Well-Architected questions closed	5 of 5
One-time charges this week	\$3,164
Cost avoidance at the December volume, not counted	\$140,000 to \$230,000 a year
Measured savings	none until week six

What the sponsor needs to do

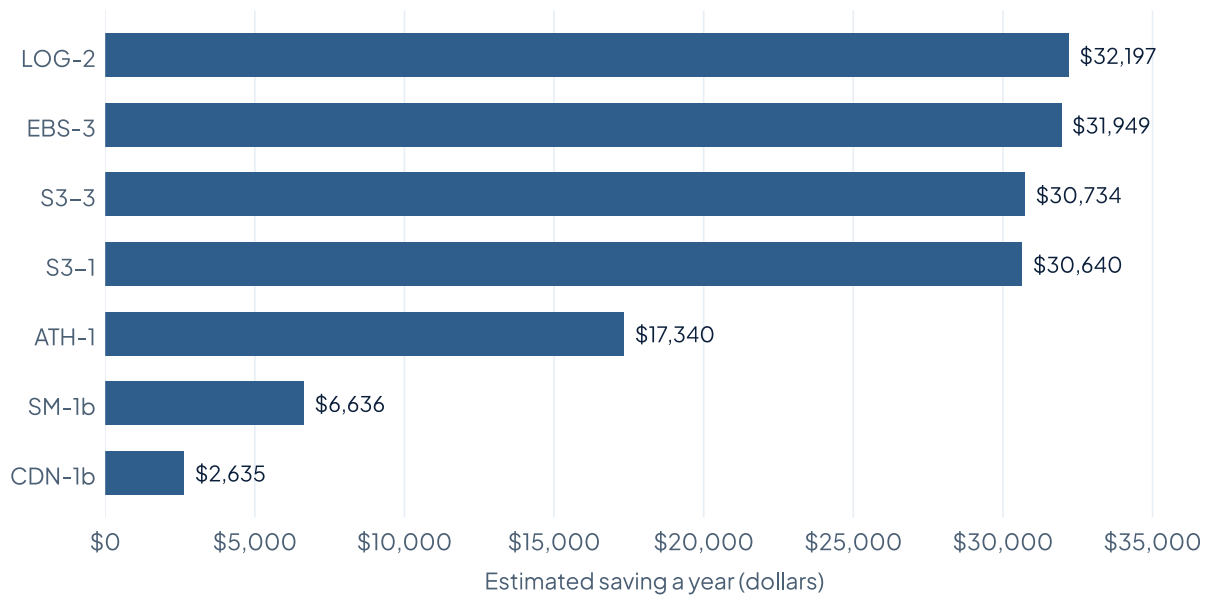
Dev Raman: nothing is needed to keep this week’s savings, and one ask is open. Please find out from counsel when the retention schedule will be signed, because S3-2, the largest change in the register at \$120,787 a year, is deferred to May 20 and waits on it, and so do the C1.2 controls in the SOC 2 map. CACHE-2, Redis OSS to Valkey, is deferred to May 19 until the client library is tested against Valkey; that test is Kofi Brandvold’s team’s.

What went in

Five kinds of change went in. The largest by value is architecture, \$50,615 a year across three changes. LOG-2 (Shorter application log retention, counsel’s copy first) is the largest single change at \$32,197.

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 5
Architecture	S3-1, ATH-1, CDN-1b	3	\$50,615	33.3%
Logging	LOG-2	1	\$32,197	21.2%
Waste	EBS-3	1	\$31,949	21.0%
Storage	S3-3	1	\$30,734	20.2%
Rightsizing	SM-1b	1	\$6,636	4.4%
Total		7	\$152,131	100.0%



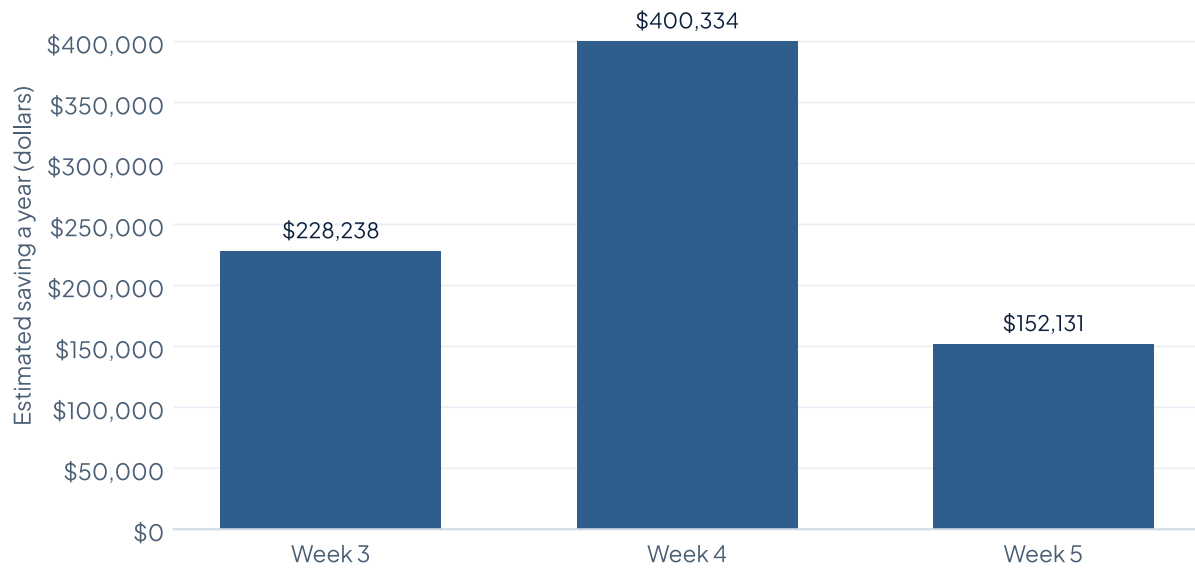
Estimated saving a year for each change in place in week 5.

When each change went in. Wednesday, May 6, is the last day a change can go in and still have seven full days on the Cost and Usage Report before it is read.

Day	In place	Changes	Estimate, a year
Monday, May 4	SM-1b, EBS-3	2	\$38,585
Tuesday, May 5	LOG-2, S3-3	2	\$62,931
Wednesday, May 6	S3-1, ATH-1, CDN-1b	3	\$50,615
Total		7	\$152,131

The three weeks of changes, cumulative. 31 changes are in place at an estimated \$780,703 a year, 37.0% of the April bill annualized.

Week	Dates	Changes in place	Estimate, a year
Week 3	April 20 to April 24	13	\$228,238
Week 4	April 27 to May 1	11	\$400,334
Week 5	May 4 to May 8	7	\$152,131
Total		31	\$780,703



Estimated saving a year of the changes put in place each week, at list price. Not measured.

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 5, in the order it went in. Estimates are annualized at list price; nothing is measured.

Change	Before	After	Way back	Estimate, a year
SM-1b The embedding endpoint scales on its backlog. tp-prod; Yuki Sorensen; Monday, May 4.	The embedding endpoint as SM-1a leaves it: a fixed instance count of the smaller type.	Embedding endpoint scales on its backlog, one to four instances	Remove the scaling policy and set the instance count back to its previous fixed size.	\$6,636
EBS-3 Snapshot chains past their keep period deleted. tp-prod; Kofi Brandvold; Monday, May 4.	Snapshots from a cron job that never deletes: 65 TB billed, at \$0.05 a GB-month.	The cron's snapshot chains past 35 days deleted; AWS Backup governs from here	No way back for a deleted snapshot. Recovery points come from AWS Backup, and its restore test is the proof that a restore from the vault works.	\$31,949
LOG-2 Shorter application log retention, counsel's copy first. tp-prod; Sasha Ilyin; Tuesday, May 5.	Log storage, never expiring: 89.8 TB stored, at \$0.03 a GB-month.	Application log groups kept 30 days; January to March exported for counsel first and kept in S3 Glacier Deep Archive	No way back for an event that has aged out. Counsel's copy of January to March is in Deep Archive, and a restore takes hours; the retention period can be raised again for new events.	\$32,197
S3-3 Old renditions to Glacier Instant Retrieval. tp-prod; Amara Delacroix; Tuesday, May 5.	S3 Standard: originals, renditions, raw telemetry JSON, exports, and artifacts: 1,087 TB (709 originals, 207 renditions, 151 raw telemetry, 20 exports and artifacts), at the published tiers.	Renditions older than 60 days to S3 Glacier Instant Retrieval (priced against Standard-IA and Intelligent-Tiering on the same inventory)	Remove the rule. Objects already in Glacier Instant Retrieval stay there until a copy-in-place or a second rule moves them back, reads cost the retrieval charge, and an object moved out early bills the rest of its minimum storage period.	\$30,734

Change	Before	After	Way back	Estimate, a year
S3-1 Telemetry to Parquet at ingest. tp-prod; Lior Haddad; Wednesday, May 6.	S3 Standard: originals, renditions, raw telemetry JSON, exports, and artifacts: 1,087 TB (709 originals, 207 renditions, 151 raw telemetry, 20 exports and artifacts), at the published tiers.	Telemetry to Parquet at ingest; 151 TB of raw JSON converted, then moved to S3 Glacier Deep Archive	Restore the raw JSON from Deep Archive when needed (hours, at a retrieval charge). The Parquet is a lossless conversion of the same events, so a restore is not needed to query them.	\$30,640
ATH-1 Athena on partitioned Parquet. tp-prod; Lior Haddad; Wednesday, May 6.	Athena on raw JSON: 330 TB scanned, at \$5.00 a TB.	Athena on partitioned Parquet	Point the queries back at the JSON table, which holds only what has not yet moved to Deep Archive. There is no reason to: Parquet holds the same events.	\$17,340
CDN-1b One rendition fewer to transcode. tp-prod; Kofi Brandvold; Wednesday, May 6.	The transcoder fleet as EC2-5 leaves it, producing every rendition in the ladder.	One rendition fewer to transcode	Restore the rendition in the ladder configuration. New uploads get it again; nothing is regenerated for earlier ones.	\$2,635
Total		7 changes		\$152,131

Telemetry: Parquet and Athena

Tidepost’s apps send about 1,400 events per active user per day, and until this week every one landed as a raw JSON object in S3 Standard and was scanned, whole, by Athena. S3–1 (Lior Haddad, Wednesday, May 6) changes both ends: Firehose converts the stream to Parquet at ingest, and a Glue job converts what was already there; ATH–1 (the same day) points the analysts’ queries at the partitioned Parquet table. The Well-Architected question PERFO3 (“how do you store, manage, and access data in your workload?”) closes with these two.

The telemetry path before and after; the estimates are \$30,640 a year for storage and \$17,340 for queries. Athena’s query history showed that every query reads one to seven days of data, which is why the older raw JSON can sit in Deep Archive.

Stage	Before	After
New events	Raw JSON objects, S3 Standard	Parquet at ingest through a Glue table, S3 Standard
Events already stored	Raw JSON in S3 Standard	Converted, verified, then the raw JSON moved to S3 Glacier Deep Archive
Analyst queries	Athena scans whole days of JSON for a few columns	Athena reads the columns it names from the days it names

What it cost to do. S3–1 carries a one-time charge of \$461 (Glue conversion and the Deep Archive transition requests), at list. It is listed here and in the week–six report beside the result and never netted into the annual savings. **The estimate** for S3–1 is 151 TB at \$0.021 to 18.9 TB of Parquet, the JSON at \$0.00099, plus Firehose format conversion; Deep Archive bills a minimum storage period, so the raw JSON is parked there for good and not rotated. **Why it was safe:** Athena query history: every query reads one to seven days; Lior Haddad’s team compares row counts for each converted partition before the raw JSON moves, and reruns any partition that fails.

Storage tiers: renditions now, originals on May 20

Renditions to Glacier Instant Retrieval, priced against Standard–IA and Intelligent–Tiering

S3–3 moves the renditions older than 60 days to S3 Glacier Instant Retrieval, which still serves CloudFront’s fetch of an old segment in milliseconds. Three storage classes could hold them, and all three were priced on the same input: the renditions prefix in the S3 Inventory of April 6, by object size and by age. Renditions are written once, so an object’s age is its last–modified date.

The S3 Inventory of April 6, by object size. The smallest band is playlists, init segments, and poster frames.

Renditions older than 60 days, by size	Objects	TB
Under 128 KB	22,400,000	1.3
128 KB to 1 MB	41,600,000	24.6
1 MB to 4 MB	82,200,000	128.4
Over 4 MB	4,200,000	23.7
Total	150,400,000	178.0

The 22,400,000 objects under 128 KB (1.3 TB) stay in S3 Standard whichever option is chosen: a lifecycle rule skips an object that small by default, and Intelligent-Tiering neither monitors nor tiers one. The other 128,000,000 objects, 176.7 TB averaging 1.4 MB, are what each option moves. Today they cost \$3,800 a month in S3 Standard at \$0.021 a GB-month.

The objects of 128 KB and up by age, with the bytes CloudFront's origin fetches read back from each band in a month. Renditions older than 60 days serve 2% of views.

Those objects, by age	Objects	TB	Read back a month
60 to 90 days	13,300,000	18.4	5.0 TB
90 to 150 days	24,900,000	34.3	3.0 TB
Over 150 days	89,800,000	124.0	2.0 TB
Total	128,000,000	176.7	10.0 TB

- **Standard-IA** at 60 days: \$0.0125 a GB-month and \$0.01 a GB read back.
- **Intelligent-Tiering**, all three tiers: an object that enters at 60 days old stays in Frequent Access, at the Standard price, for 30 days; it moves to Infrequent Access (\$0.0125) after 30 days without a read and to Archive Instant Access (\$0.004) after 90, and a read moves it back to Frequent Access. There is no retrieval charge, and monitoring costs \$0.0025 per 1,000 objects a month.
- **Glacier Instant Retrieval** at 60 days: \$0.004 a GB-month and \$0.03 a GB read back.

Each option also pays the lifecycle transition requests for each month's 10,400,000 objects that turn 60 days old: \$0.01 per 1,000 into Standard-IA or Intelligent-Tiering, \$0.02 into Glacier Instant Retrieval.

The three options for the old renditions, priced on the S3 Inventory. Intelligent-Tiering's storage is 23.4 TB in Frequent Access (the youngest band, 60 to 90 days old, and a month's reads of the older ones), 35.3 TB

in Infrequent Access, and 118.0 TB in Archive Instant Access. The Glacier Instant Retrieval figure is the estimate in the register.

A month, at April usage	Standard-IA	Intelligent-Tiering	Glacier Instant Retrieval (S3-3)
Storage	\$2,262	\$1,438	\$724
Reads (10 TB back)	\$102	none	\$307
Monitoring	none	\$320	none
Transition requests, each month's cohort	\$104	\$104	\$208
Cost a month	\$2,468	\$1,862	\$1,239
Saving a month against \$3,800 in Standard	\$1,332	\$1,938	\$2,561
Saving a year	\$15,981	\$23,253	\$30,734

Intelligent-Tiering ends up at the prices the other two charge: its Infrequent Access tier is Standard-IA's price, and its Archive Instant Access tier is Glacier Instant Retrieval's. It gets there later and pays to get there: \$320 a month to monitor 128,000,000 objects, the youngest band at the Standard price for its first 30 days, and every byte that is read back in Frequent Access for another 30. Its one advantage is that reads are free, worth \$307 a month against Glacier Instant Retrieval at the read-back the logs show. It saves \$7,272 a year more than Standard-IA and \$7,481 less than Glacier Instant Retrieval. It would also have shown nothing in the week after the change: every object enters at Frequent Access and moves only after 30 days without a read, so the measurement would have read its monitoring fee and none of its saving.

Glacier Instant Retrieval saves \$14,753 a year more than Standard-IA, and it would stop beating Standard-IA only if more than **70 TB a month** were read back, 7.0 times the estimate. At the 15 TB the after-window read back, the order holds: \$28,891 a year for Glacier Instant Retrieval, \$21,164 for Intelligent-Tiering, and \$15,367 for Standard-IA. Glacier Instant Retrieval bills a 90-day minimum, so a takedown that removes a rendition within 90 days of its move bills the rest of those days. S3-3 carries a one-time charge of \$2,560 for the 128,000,000 transition requests, at \$0.02 per 1,000.

Originals to Glacier Instant Retrieval: S3-2, deferred to May 20

S3-2 is the largest change in the register, \$120,787 a year, and it is not in place. Deferred to May 20: approved, and the rule waits for counsel to sign the retention schedule. The rule moves originals older than 90 days to S3 Glacier Instant Retrieval, which still serves a takedown or legal

request in milliseconds. Counsel’s retention schedule is the input: the rule must never move or expire an original that counsel’s schedule says to keep in Standard, and the schedule is not signed. Its one-time charge is \$192 when it lands.

The same schedule is a Trust Services Criterion C1.2 control in the SOC 2 technical map: retention schedule for user content, written with counsel; deletion pipeline for removed posts’ media, across originals, renditions, and backups. Those controls begin operating on May 20, after the observation window opened on April 1; the map says so and the auditor decides what that means.

Logs: LOG-2, and the audit logs it leaves alone

LOG-2 shortens how long CloudWatch keeps the application log groups, after counsel has a copy of January to March. It touches only application logs. The audit logs, the organization trail and the AWS Config history in tp-log-archive, are a different set with a different rule.

Which logs LOG-2 changes and which it does not. The auditor tests the audit logs’ retention; this change does not shorten it.

Log	Where	Retention	Changed by LOG-2
Application log groups	CloudWatch Logs, tp-prod	Application log groups kept 30 days; January to March exported for counsel first and kept in S3 Glacier Deep Archive	Yes
January to March of those log groups	S3 Glacier Deep Archive, for counsel	Held for counsel	New: exported before anything expired
The organization trail and AWS Config history	S3, tp-log-archive	Audit logs live 400 days in tp-log-archive (the security engagement), so only application logs change	No

The order mattered. Sasha Ilyin confirmed that counsel’s copy had landed in Deep Archive on Tuesday, May 5, and only then did the retention period drop. LOG-2 is the second step after LOG-1 (the log level, week 3): INFO writes less, and shorter retention keeps less of it. **The estimate** is 89.8 TB stored to 1.8 TB at \$0.03, plus 20 TB in Deep Archive at \$0.00099.

Snapshots: EBS-3 and AWS Backup

A cron job had taken EBS snapshots on its own schedule and never deleted one. EBS-3 deletes the chains past their keep period and retires the cron job in the same pull request; from here AWS Backup in tp-backup, with its vault lock in compliance mode, is the only schedule and the only retention. Deleting a snapshot frees only the blocks no other snapshot references, which is why the estimate counts the chains' unique blocks, read from the EBS direct APIs, and not the volumes' size. The restore test that proves a recovery point comes back from the vault was done on March 24, in the Well-Architected Remediation, and it is the evidence that deleting the old chains does not remove the only way back.

The SageMaker Savings Plan: SP-2

SP-2 was approved at the April 17 readout and declined on Tuesday, May 5. Approved at the readout, then declined on May 5 by Gwen Tallis: she wants the Inferentia2 move first. It is a one-year, No Upfront plan at \$4.14 an hour, 80% of the hosting floor's plan cost (\$6.78 an hour at On-Demand, \$5.18 at the plan rate), covering \$5.42 of On-Demand hosting. It would have saved an estimated \$11,192 a year, and it would have committed Tidepost to \$36,266 over the term.

The decision is a good one on the facts. The floor SP-2 is sized on still holds the moderation classifier on GPU instances; SM-3 moves it to Inferentia2 and is deferred to July 13 because it needs a Neuron compile and a precision check first. A plan bought before the move would be sized on the wrong floor. Gwen Tallis can buy it after the move, on the floor the move leaves. SP-2 stays in the register as approved, then declined; nothing is measured for it, and the week-six report shows it on its own line and never inside a measured total.

Guardrails

A change that saves money today does nothing about the next engineer who creates the same waste. The Optimization hands over seven guardrails. It builds five of them in week 5 (GR-1, GR-2, GR-3, GR-4, GR-6); the other two are inherited with their original dates: GR-5, from the Well-Architected Remediation, which closed COST02 with it on March 13; GR-7, from the Foundation, whose tag policy was attached on February 10 and whose cost allocation tags the Well-Architected Remediation activated on March 11 for COST03. The Optimization adds those two to its handover and its monthly review and builds nothing new for them. Each guardrail has one

owner, who is the person who will be asked if it fails. Four prevent the waste at creation (GR-1, GR-3, GR-4, GR-6) and three detect it soon after (GR-2, GR-5, GR-7); the table says which, and what each cannot see.

The seven guardrails, one owner each: five built in week 5 and two inherited (GR-5 and GR-7). "Keeps in place" names the week 3 to 5 changes whose waste each would stop from returning.

Guardrail	Owner	How it works	Prevents or detects	What it does not see	Keeps in place
GR-1 Log level and retention defaults on every new log group (INFO, 30 days)	Marcus Oyelaran	A CDK aspect fails the synth without them	Prevents	It sees only log groups defined in tidepost-cdk; a log group created by hand is invisible to a CDK aspect.	LOG-1, LOG-2
GR-2 The lifecycle standard on every new bucket	Kofi Brandvold	An AWS Config rule flags a bucket without one within an hour	Detects	It detects, it does not block: a bucket can exist without a lifecycle rule until the Config rule flags it within the hour.	S3-1, S3-3
GR-3 Notebook auto-stop by default in tp-sandbox-ml	Yuki Sorensen	A lifecycle configuration attached by Account Factory	Prevents	It covers notebooks vended through Account Factory in tp-sandbox-ml; a notebook created anywhere else is outside it.	SM-2
GR-4 A scaling policy required on every SageMaker endpoint	Yuki Sorensen	The deploy pipeline refuses an endpoint without one	Prevents	It gates the deploy pipeline; an endpoint created by hand in the console never passes the gate.	SM-1a, SM-1b
GR-5 Budgets and anomaly alerts per account	Gwen Tallis	AWS Budgets and Cost Anomaly Detection, to Gwen Tallis and Kofi Brandvold	Detects	It alerts two people and blocks nothing; a spike is seen when the anomaly detector finds it, not when it starts.	LOG-3, ATH-1
GR-6 A Kinesis retention cap of seven days	Lior Haddad	A CDK aspect fails the synth above 168 hours, and an EventBridge rule alerts the data team on any IncreaseStreamRetentionPeriod call	Prevents and detects	The cap fails the synth, and the EventBridge rule alerts on a manual change; the alert reaches the data team after the call, not before.	KIN-1, KIN-2

Guardrail	Owner	How it works	Prevents or detects	What it does not see	Keeps in place
GR-7 The Foundation's tag policy, enforced on the cost allocation tags	Kofi Brandvold	The tag policy's compliance report, read in the monthly review	Detects	It is a compliance report read in the monthly review, so a resource without the cost allocation tags is found at the next review, not at creation.	The tags every report in this engagement reads

GR-7 deserves a line of its own, because it is the one that makes the others measurable. The Foundation's tag policy requires the keys team, environment, and service on the resources that matter, and the cost allocation tags were activated on March 11 in the Well-Architected Remediation (COST03). Kofi Brandvold reads the tag policy's compliance report in the monthly review. A resource without the tags cannot be allocated, and a cost that cannot be allocated cannot be owned.

Cost avoidance, not counted

This is cost avoidance, not savings, and it is not counted

Everything above is a saving against what Tidepost pays today. What follows is different: an estimate of what the old configurations would have cost by December if usage grows as planned. It is a model with assumptions, it is labeled cost avoidance, and it is not counted toward the engagement's savings, the gate, or the fee. Gwen Tallis's CloudFront private pricing negotiation is named in the roadmap and is not counted either.

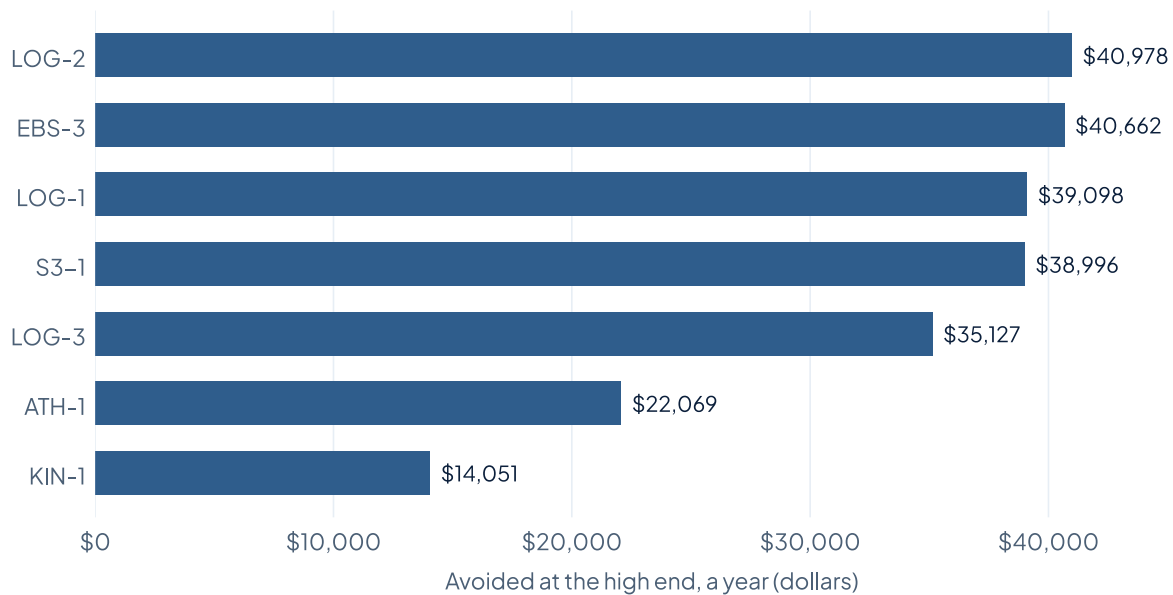
The setup. Usage on the April 6 bill is 660,000 monthly users: the 600,000 of January plus the growth to April that the bill carries. Tidepost plans 1,500,000 for December. That is 2.27 times today, an increase of 127%. A change's yearly estimate is the waste it removes at today's usage. If the old configuration had stayed while usage grew, that waste would have grown with it, and the growth is what the change avoids.

The levers. Seven of the changes remove a cost that rises with users, and each is in place before December (LOG-1, LOG-2, LOG-3, KIN-1, S3-1, ATH-1, EBS-3; every one by May 6). The changes that scale on demand already, such as the API tier and the transcoders, are left out: they size themselves, so there is nothing avoided. So are the notebooks, which scale with the machine-learning team and not with users.

The levers behind the cost avoidance range, not counted. Each lever's yearly estimate times the growth (127%), at 60% and at 100% of linear. The total rounds to the nearest \$10,000.

Lever	What scales with users	Held by	Estimate today, a year	Avoided at 60% of linear	Avoided at 100%
LOG-1 INFO in production, not DEBUG	Logs written per request	GR-1	\$30,720	\$23,459	\$39,098
LOG-2 Shorter application log retention, counsel's copy first	Log events kept, which accumulate with ingestion	GR-1	\$32,197	\$24,587	\$40,978
LOG-3 Custom metrics without the per-post dimension	Metric series, which multiply with posts	GR-5	\$27,600	\$21,076	\$35,127
KIN-1 Kinesis retention cut back	Records held past the point anything reads them	GR-6	\$11,040	\$8,431	\$14,051
S3-1 Telemetry to Parquet at ingest	Raw telemetry bytes, which arrive with every event	GR-2	\$30,640	\$23,398	\$38,996
ATH-1 Athena on partitioned Parquet	Bytes scanned, which grow with the data queried	GR-5	\$17,340	\$13,241	\$22,069
EBS-3 Snapshot chains past their keep period deleted	Snapshot blocks, which grow with the data on the volumes	AWS Backup's lifecycle	\$31,949	\$24,397	\$40,662
Total			\$181,486	\$138,589	\$230,982

The two ends. At the high end, every lever grows in step with users. At the low end, 60% of that growth appears, because volume per user can fall as sampling and caching improve. Those two percentages are Vantalect's assumption, not a measurement and not a fact from Tidepost's data, and they are the part of the model most worth challenging. Rounded to the nearest \$10,000, the range is **\$140,000 to \$230,000 a year** at the December volume.



Cost avoided by lever at the high end, a year. Cost avoidance, not counted.

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 are the same as in week 3's report: the pull request, the pipeline run, the CloudFormation drift report, the API call in the organization trail, and the change window.

The repository, date, and window class for every change in week 5. The guardrails' own code is in tidepost-cdk; its pull requests are the evidence that each guardrail was built.

Change	Repository of the pull request	Date	Window	Approved by
SM-1b	tidepost-cdk	Monday, May 4	Standard	Yuki Sorensen, at the April 17 readout
EBS-3	tidepost-cdk	Monday, May 4	Standard	Kofi Brandvold, at the April 17 readout
LOG-2	tidepost-cdk	Tuesday, May 5	Standard	Sasha Ilyin, at the April 17 readout
S3-3	tidepost-cdk	Tuesday, May 5	Standard	Amara Delacroix, at the April 17 readout
S3-1	tidepost-cdk	Wednesday, May 6	Standard	Lior Haddad, at the April 17 readout
ATH-1	tidepost-cdk	Wednesday, May 6	Standard	Lior Haddad, at the April 17 readout
CDN-1b	tidepost-cdk	Wednesday, May 6	Standard	Kofi Brandvold, 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-1b: The embedding endpoint scales on its backlog

Owner: Yuki Sorensen. **Account:** tp-prod. **In place:** Monday, May 4. **Estimate:** \$6,636 a year (\$553 a month). **What changed:** Embedding endpoint scales on its backlog, one to four instances.

How it was made. Application Auto Scaling scales the asynchronous endpoint on its backlog per instance, between a floor of one instance and a ceiling of four. The queue is empty most of the day, so the endpoint idles at the floor and grows when a backlog builds. An asynchronous endpoint can scale to zero, and this one does not: a cold start, an instance launch and the model load, takes about six minutes, and a new post must be embedded within two minutes to reach the ranking feed, so one instance stays warm. **The estimate** is $(2 - 1.5) \text{ instances} \times \$1.515 \times 730 \text{ hours}$.

Why it was safe: The asynchronous queue's backlog: empty 70% of the day; the floor is one, not

zero: one instance stays warm because a cold start (an instance launch and the model load) takes about six minutes, and new posts must be embedded within two minutes to reach the ranking feed.

Way back. Remove the scaling policy and set the instance count back to its previous fixed size.

What the owner watches. Yuki Sorensen's team watches the age of the oldest request in the queue. Week five makes a scaling policy mandatory for every SageMaker endpoint (GR-4).

Next.

- **GR-4** (built in week 5): requires a scaling policy on every SageMaker endpoint.
- **SM-3** (deferred to Monday, July 13): moves the moderation classifier to Inferentia2.

EBS-3: Snapshot chains past their keep period deleted

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Monday, May 4. **Estimate:** \$31,949 a year (\$2,662 a month). **What changed:** The cron's snapshot chains past 35 days deleted; AWS Backup governs from here.

How it was made. The cron job that took the snapshots is retired in the pull request, and a reviewed script deletes the chains it left past the keep period. The first run only listed every snapshot with its unique blocks. AWS Backup in tp-backup, with its vault lock in compliance mode, is the only schedule and retention from here. **The estimate** is 52 TB of unique blocks at \$0.05.

Why it was safe: The snapshot chains' unique blocks, from the EBS direct APIs.

Way back. No way back for a deleted snapshot. Recovery points come from AWS Backup, and its restore test is the proof that a restore from the vault works. **What the owner watches.** Kofi Brandvold's team checks that every volume with the backup tag has a recovery point in the vault after the first scheduled run.

LOG-2: Shorter application log retention, counsel's copy first

Owner: Sasha Ilyin. **Account:** tp-prod. **In place:** Tuesday, May 5. **Estimate:** \$32,197 a year (\$2,683 a month). **What changed:** Application log groups kept 30 days; January to March exported for counsel first and kept in S3 Glacier Deep Archive.

How it was made. First, CloudWatch Logs export tasks write January to March of the application log groups to S3 Standard under a counsel prefix in tp-log-archive, and a lifecycle rule moves each object to S3 Glacier Deep Archive the next day; Sasha Ilyin confirms the copy landed. Only then does the retention period on the application log groups drop. CloudWatch removes expired events over about three days, so the storage line falls over the first days of the week after. The organization trail and AWS Config history in tp-log-archive, and the flow logs, are separate: they

keep the longer retention the security engagement set. **The estimate** is 89.8 TB stored to 1.8 TB at \$0.03, plus 20 TB in Deep Archive at \$0.00099. **Why it was safe:** Audit logs live 400 days in tp-log-archive (the security engagement), so only application logs change.

Way back. No way back for an event that has aged out. Counsel's copy of January to March is in Deep Archive, and a restore takes hours; the retention period can be raised again for new events.

What the owner watches. Sasha Ilyin keeps the export's manifest with the evidence index. The auditor tests the audit logs' retention, and this change does not shorten it. **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.

- **GR-1** (built in week 5): sets the retention default on every new log group.

S3-3: Old renditions to Glacier Instant Retrieval

Owner: Amara Delacroix. **Account:** tp-prod. **In place:** Tuesday, May 5. **Estimate:** \$30,734 a year (\$2,561 a month). **What changed:** Renditions older than 60 days to S3 Glacier Instant Retrieval (priced against Standard-IA and Intelligent-Tiering on the same inventory).

How it was made. A lifecycle rule on the renditions bucket transitions objects to S3 Glacier Instant Retrieval once they are older than the age in the title; objects below the minimum billable size stay in Standard, as lifecycle rules leave them by default. The rule is the one the lifecycle standard (GR-2) will require on every new bucket. **The estimate** is 176.7 TB of 128 KB and up from \$0.021 to \$0.004, plus 10 TB a month read back at \$0.03 and the transition requests for each month's 10,400,000 objects that turn 60 days old. **Why it was safe:** S3 Inventory of April 6 by size and age; CloudFront logs: renditions older than 60 days serve 2% of views and 10 TB a month is read back.

Way back. Remove the rule. Objects already in Glacier Instant Retrieval stay there until a copy-in-place or a second rule moves them back, reads cost the retrieval charge, and an object moved out early bills the rest of its minimum storage period. **What the owner watches.** Amara Delacroix's team and Marcus Oyelaran watch the share of views that read an old rendition; the retrieval charge grows if a story revives old videos. **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.

- **S3-2** (deferred to Wednesday, May 20): originals follow once counsel signs the retention schedule.
- **GR-2** (built in week 5): puts a lifecycle standard on every new bucket.

S3-1: Telemetry to Parquet at ingest

Owner: Lior Haddad. **Account:** tp-prod. **In place:** Wednesday, May 6. **Estimate:** \$30,640 a year (\$2,553 a month). **What changed:** Telemetry to Parquet at ingest; 151 TB of raw JSON converted, then moved to S3 Glacier Deep Archive.

How it was made. Firehose converts the telemetry stream to Parquet at ingest, with a Glue table defining the schema. A Glue job converts the raw JSON already in the bucket, and once a partition's conversion is verified the raw JSON moves to S3 Glacier Deep Archive, billed for a minimum storage period there. Nothing is deleted. **The estimate** is 151 TB at \$0.021 to 18.9 TB of Parquet, the JSON at \$0.00099, plus Firehose format conversion. **Why it was safe:** Athena query history: every query reads one to seven days.

Way back. Restore the raw JSON from Deep Archive when needed (hours, at a retrieval charge). The Parquet is a lossless conversion of the same events, so a restore is not needed to query them. **What the owner watches.** Lior Haddad's team compares row counts for each converted partition against the JSON before the move, and reruns any partition that fails. **Well-Architected.** This change is on the list for PERFO3 (closes May 6, when S3-1 lands) and SUSO4 (closes May 6, when S3-1 lands), accepted in writing and taken up by the cost engagement.

Next.

- **ATH-1** (in place Wednesday, May 6): the queries move to the Parquet table.
- **GR-2** (built in week 5): puts a lifecycle standard on every new bucket.

ATH-1: Athena on partitioned Parquet

Owner: Lior Haddad. **Account:** tp-prod. **In place:** Wednesday, May 6. **Estimate:** \$17,340 a year (\$1,445 a month). **What changed:** Athena on partitioned Parquet.

How it was made. The analysts' queries point at the partitioned Parquet table. A query that names a day and a few columns reads only those, where it used to scan whole days of JSON. **The estimate** is From 330 TB to 41 TB scanned a month at \$5.00 a TB. **Why it was safe:** Each query scans whole days of JSON for a few columns.

Way back. Point the queries back at the JSON table, which holds only what has not yet moved to Deep Archive. There is no reason to: Parquet holds the same events. **What the owner watches.** Lior Haddad's team watches the bytes scanned each day in Athena's query history. **Well-Architected.** This change is on the list for PERFO3 (closes May 6, when S3-1 lands), accepted in writing and taken up by the cost engagement.

Next.

- **GR-5** (in place since March 13, from the Well-Architected Remediation): budgets and anomaly alerts would notice scanned bytes that climb again.

CDN-1b: One rendition fewer to transcode

Owner: Kofi Brandvold. **Account:** tp-prod. **In place:** Wednesday, May 6. **Estimate:** \$2,635 a year (\$220 a month). **What changed:** One rendition fewer to transcode.

How it was made. With CDN-1 in the apps, the transcoder pipeline stops producing the 1080p rendition for new vertical uploads. Existing videos keep the renditions they have. **The estimate** is 0.4 g4dn.2xlarge on average at \$0.752 x 730 hours (priced after EC2-5). **Why it was safe:** The 1080p rendition takes 13% of transcoding time.

Way back. Restore the rendition in the ladder configuration. New uploads get it again; nothing is regenerated for earlier ones. **What the owner watches.** Kofi Brandvold's team watches the transcoding queue and playback errors on new uploads.

What cannot be undone

Two changes cannot be fully reversed

EBS-3 (Snapshot chains past their keep period deleted): No way back for a deleted snapshot. Recovery points come from AWS Backup, and its restore test is the proof that a restore from the vault works. **LOG-2** (Shorter application log retention, counsel's copy first): No way back for an event that has aged out. Counsel's copy of January to March is in Deep Archive, and a restore takes hours; the retention period can be raised again for new events. The other changes this week reverse with a pull request or the step beside them. The raw telemetry JSON is not lost by S3-1: it is in Deep Archive and restorable in hours.

The Well-Architected questions

All five of the Well-Architected questions accepted in writing now have their fixes in place. Two closed this week, on May 6: PERF03 and SUS04. Tidepost saves its answers in the Well-Architected Tool at the May 14 milestone, and the Tool's count of high-risk issues is the Tool's and Tidepost's to report.

The five questions accepted in writing. A question's fix is in place when the last change on its list is in place.

Question	Signed by	Changes on its list	Last change in	Fix in place since
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	SM-1a	April 27
SUS02 How do you align cloud resources to your demand?	Yuki Sorensen, March 11	SM-1a, SM-2	SM-1a	April 27
COST07 How do you use pricing models to reduce cost?	Gwen Tallis, March 10	SP-1	SP-1	April 30
PERF03 How do you store, manage, and access data in your workload?	Lior Haddad, March 9	S3-1, ATH-1	S3-1	May 6
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	S3-1	May 6

Still ahead, and how week six measures

Five changes are deferred, each with its reason and its date. None is dropped, and none is in any figure above.

The deferred changes with the dates the facts give.

Change	Date	Owner	Why it waits	Estimate, a year
CACHE-2 Redis OSS to Valkey on the same nodes	Tuesday, May 19	Kofi Brandvold	Deferred to May 19: the client library is tested against Valkey first	\$5,754
S3-2 Originals to S3 Glacier Instant Retrieval	Wednesday, May 20	Amara Delacroix	Deferred to May 20: approved, and the rule waits for counsel to sign the retention schedule	\$120,787
EC2-1a Staging rebuilt smaller in tp-staging	Monday, June 1	Kofi Brandvold	Deferred to June 1: two platform engineers, and the rebuild uses the Remediation's CDK	\$55,468
EC2-1b Staging on weekdays only	Monday, June 8	Kofi Brandvold	Deferred to June 8: it follows EC2-1a	\$11,218
SM-3 The moderation classifier on Inferentia2	Monday, July 13	Yuki Sorensen	Deferred to July 13: it needs a Neuron compile and a precision check against the labeled March set	\$7,323

How the changes are measured. The week-six report, on May 15, reads each change on the Cost and Usage Report, which runs through May 13. Non-commitment changes are measured on the public On-Demand cost of the changed resources (including the On-Demand equivalent where a plan covered the usage), seven days before against seven days after, the change day excluded, then annualized. The Savings Plan is measured by its ARN: the On-Demand cost of the usage it covered, less its fee. S3 changes are measured by prefix and storage class from S3 Inventory and Storage Lens. The rendition ladder is normalized to bytes per view, so that growth in views is not counted as a saving. Business Support+ is billed monthly, appears in no daily series, and carries no tags, so it is outside every daily measurement. May 6 is the last day a change could go in and still have seven full days after it; the deferred changes are not measured in this engagement.

One-time charges. LOG-2 and S3-3 and S3-1 carry \$3,164 in one-time charges this week (\$143 on LOG-2, the export's writes to S3 Standard, its day there, and the Deep Archive transition requests; \$2,560 on S3-3, Glacier Instant Retrieval transition requests; \$461 on S3-1, Glue conversion and Deep Archive transition requests). They are shown beside the result in week six and never netted into the annual savings.

© 2026 Vantalect, LLC. Sample deliverable; Tidepost, Inc. is fictional.

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.