



AWS WELL-ARCHITECTED FRAMEWORK REVIEW · SAMPLE DELIVERABLE

# Review Sponsor Readout and Milestone One

Week 2 of 6: Sponsor readout · prepared for Tidepost, Inc.

---

Prepared by Charles Sieg · [charles.sieg@vantalect.com](mailto:charles.sieg@vantalect.com) · [vantalect.com](https://vantalect.com)

March 6, 2026 · AWS Well-Architected Framework Review: The Remediation (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

---

- The score ..... 2
- What the score means ..... 4
- The agreed register ..... 4
- Decisions we need from you ..... 6
  - The five HRIs to accept in writing ..... 6
  - The PostgreSQL upgrade waits for the cutover ..... 7
  - The originals in a second Region..... 8
  - Decided today ..... 8
- What the Remediation adds to the bill ..... 8
- The four weeks ..... 9
- The gate and the fee ..... 10

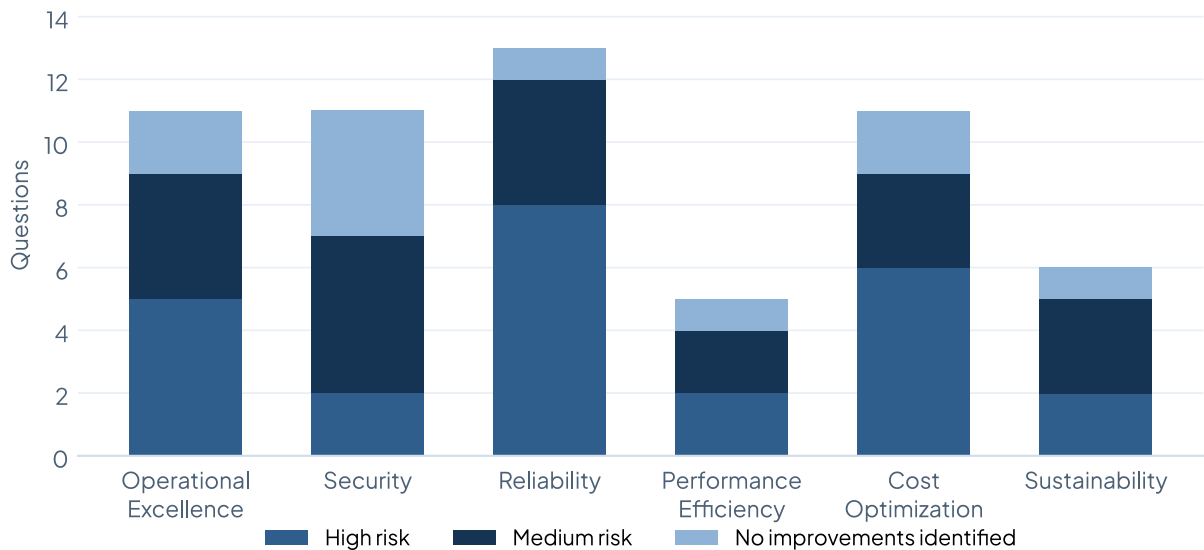
## The score

---

Across the 57 questions of the AWS Well-Architected Framework lens, the Brief has 25 high-risk issues, 21 medium-risk issues, and 11 questions with no improvements identified. Dev Raman: this is the Review’s readout. Every answer was checked against tp-prod’s live configuration, its telemetry, its bill, and the tidepost-cdk repository, and it is recorded in Tidepost’s own AWS Well-Architected Tool, where we saved it today as milestone one, the baseline the Remediation is measured against.

Milestone one, saved in the AWS Well-Architected Tool on March 6, 2026. Each question counts once, at the risk the Tool gives its answer.

| Pillar                 | Questions | High risk | Medium risk | No improvements identified |
|------------------------|-----------|-----------|-------------|----------------------------|
| Operational Excellence | 11        | 5         | 4           | 2                          |
| Security               | 11        | 2         | 5           | 4                          |
| Reliability            | 13        | 8         | 4           | 1                          |
| Performance Efficiency | 5         | 2         | 2           | 1                          |
| Cost Optimization      | 11        | 6         | 3           | 2                          |
| Sustainability         | 6         | 2         | 3           | 1                          |
| All six pillars        | 57        | 25        | 21          | 11                         |



Risk by pillar at milestone one, March 6, 2026.

Reliability carries the most high risk, 8 of its 13 questions, and Security the least, because the security engagement closed most of what that pillar asks about before the Review began. The Machine Learning Lens informed the questions about the ranking and moderation endpoints; it is context only, and none of its questions is answered or counted.

## What the score means

Four facts produce most of the 25. Each is one sentence of evidence from the Tool.

- **One Availability Zone.** Every stateful service and every EC2 tier sits in one Availability Zone. A zone failure takes the Brief down, and nothing has ever tested what comes back (REL12).
- **No recovery plan.** No recovery objectives are defined, there is nothing outside us-east-1, and recovery has never been tested.
- **Half the estate is not in code.** About half of the estate was built in the console: the moderation pipeline, the data stores, and the network. Application code is in version control, but the infrastructure under it cannot be reviewed, tested, or rebuilt from a repository.
- **Capacity that never follows demand.** GPU endpoints and notebooks run all day and night, and every developer instance runs around the clock. Nothing expires: logs, the telemetry stream, and the raw telemetry have no retention, and nothing is classified.

## The agreed register

**All 25 HRIs are agreed today: 20 close by change in weeks three to five, and 5 are to be accepted in writing, because their fix is a cost change and the cost engagement counts each change once.** Every change to the estate lands as code through Tidepost's pull requests, reviewed by the engineer who owns the service and shipped by Tidepost's pipeline; OPS07, OPS10, COST01, and SEC07 close with a runbook, a process, or a standard that the owner writes and adopts. We pair and review; we do not write to production.

The 20 HRIs that close by change, in the week each is agreed to close.

| HRI   | Question                                                                       | Fix                                                                                                         | Owner          | Week |
|-------|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|----------------|------|
| OPS04 | How do you implement observability in your workload?                           | Service-level objectives for upload-to-timeline time and playback start, with tracing and burn-rate alarms  | Kofi Brandvold | 4    |
| OPS05 | How do you reduce defects, ease remediation, and improve flow into production? | Import the console-built resources (the moderation pipeline, the data stores, and the network) into the CDK | Kofi Brandvold | 5    |
| OPS06 | How do you mitigate deployment risks?                                          | Canary deploys with automatic rollback on EKS, and a target-group swap with rollback on the EC2 tiers       | Kofi Brandvold | 5    |

| HRI   | Question                                                                                                                                  | Fix                                                                                                             | Owner           | Week |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-----------------|------|
| OPS07 | How do you know that you are ready to support a workload?                                                                                 | Runbooks, playbooks, and an on-call rotation with a primary and a secondary                                     | Kofi Brandvold  | 4    |
| OPS10 | How do you manage workload and operations events?                                                                                         | An incident process (severity levels, commander, escalation, customer message), tested in a tabletop            | Kofi Brandvold  | 5    |
| SEC07 | How do you classify your data?                                                                                                            | A data-class standard, with a data-class tag on every bucket, table, and database                               | Sasha Ilyin     | 4    |
| SEC11 | How do you incorporate and validate the security properties of applications throughout the design, development, and deployment lifecycle? | Dependency scanning and static analysis in CI, blocking on high severity                                        | Marcus Oyelaran | 4    |
| REL01 | How do you manage Service Quotas and constraints?                                                                                         | Quota alarms, and raised limits for Rekognition, SageMaker, and EC2 vCPUs                                       | Kofi Brandvold  | 3    |
| REL02 | How do you plan your network topology?                                                                                                    | Subnets and NAT gateways in three zones                                                                         | Kofi Brandvold  | 3    |
| REL07 | How do you design your workload to adapt to changes in demand?                                                                            | Target tracking on the API, Karpenter on EKS, and queue-depth scaling for the transcoders                       | Kofi Brandvold  | 4    |
| REL09 | How do you back up data?                                                                                                                  | AWS Backup to tp-backup with vault lock, DynamoDB point-in-time recovery, and a timed restore test              | Kofi Brandvold  | 5    |
| REL10 | How do you use fault isolation to protect your workload?                                                                                  | The API tier, EKS nodes, NAT, Redis, OpenSearch, and RDS across zones, proved by a zone evacuation drill        | Kofi Brandvold  | 5    |
| REL11 | How do you design your workload to withstand component failures?                                                                          | Redis replicas, OpenSearch dedicated masters, and RDS Multi-AZ failover, completed by the PostgreSQL 16 cutover | Kofi Brandvold  | 4    |
| REL12 | How do you test reliability?                                                                                                              | Playbooks, and the zone drill and the restore test run as game days                                             | Kofi Brandvold  | 5    |

| HRI    | Question                                                      | Fix                                                                                                                             | Owner          | Week |
|--------|---------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|----------------|------|
| REL13  | How do you plan for disaster recovery (DR)?                   | A recovery plan with written recovery point and recovery time objectives, and database and DynamoDB backups copied to us-west-2 | Dev Raman      | 5    |
| PERF02 | How do you select and use compute resources in your workload? | Move the ranking endpoint off GPUs if a load test shows CPU instances meet its latency budget                                   | Yuki Sorensen  | 4    |
| COST01 | How do you implement cloud financial management?              | A monthly cost review owned by Gwen Tallis and Kofi Brandvold                                                                   | Gwen Tallis    | 4    |
| COST02 | How do you govern usage?                                      | Budgets and Cost Anomaly Detection per account                                                                                  | Gwen Tallis    | 3    |
| COST03 | How do you monitor usage and cost?                            | Cost allocation tags activated (team, environment, service)                                                                     | Kofi Brandvold | 3    |
| COST04 | How do you decommission resources?                            | A decommissioning runbook, an owner tag on every resource, and a monthly unattached-resource review                             | Kofi Brandvold | 4    |

## Decisions we need from you

**Five risk acceptances, one database decision, and one copy decision, all by March 11, 2026.**  
Two more were made today.

### The five HRIs to accept in writing

The guarantee counts an HRI accepted in writing as closed. In the Tool these five stay at high risk until the cost engagement's changes land, and milestone two will show them that way.

The acceptances, each signed by the owner of the cost it carries. The cost engagement starts on April 6 and takes all five up at its readout on April 17, 2026.

| HRI    | Question                                                                                                 | The fix                                                                                                                           | Risk while it waits                                                       | Signs                                   |
|--------|----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|-----------------------------------------|
| PERF03 | How do you store, manage, and access data in your workload?                                              | Store raw telemetry in a columnar, partitioned format and point Athena at it, so a query reads only the columns and days it needs | Each Athena query keeps scanning raw JSON.                                | Lior Haddad, Head of Data               |
| COST06 | How do you meet cost targets when you select resource type, size and number?                             | Size the GPU endpoints, the notebooks, and the EC2 tiers to measured use                                                          | Oversized compute keeps running at its current size.                      | Gwen Tallis, Vice President, Finance    |
| COST07 | How do you use pricing models to reduce cost?                                                            | Commit to a Savings Plan for the steady compute once it is sized, so the commitment fits what runs                                | Everything keeps billing at On-Demand rates.                              | Gwen Tallis, Vice President, Finance    |
| SUS02  | How do you align cloud resources to your demand?                                                         | Scale the GPU endpoints and notebooks to demand instead of running them around the clock                                          | GPU endpoints and notebooks keep running around the clock.                | Yuki Sorensen, Head of Machine Learning |
| SUS04  | How do you take advantage of data management policies and patterns to support your sustainability goals? | Set retention on application logs, the telemetry stream, raw telemetry, and the originals                                         | Logs, the telemetry stream, and raw telemetry keep growing without limit. | Lior Haddad, Head of Data               |

## The PostgreSQL upgrade waits for the cutover

**We recommend upgrading production from PostgreSQL 13 to 16 in the same cutover that makes it Multi-AZ, scheduled for Sunday, March 22, and not sooner.** Standard support for PostgreSQL 13 ended on February 28, so Extended Support has billed since March 1: 32 vCPUs in production (the writer and two read replicas) and 8 in staging. Upgrading the single-AZ writer in place this week would save production's Extended Support from tomorrow to the cutover, about \$1,160, and run a major-version upgrade on a database with no standby and no tested restore. Waiting puts the upgrade on a new encrypted, private, Multi-AZ instance with the old one kept as the way back. Staging stays on 13 until Tidepost retires it, so its \$584 a month continues either way.

## The originals in a second Region

Noor Castellanos decides whether the originals bucket is copied to us-west-2. The database and DynamoDB backups go to us-west-2 in week three either way. Copying the originals means a one-time copy of the whole bucket and a second copy held in another Region for as long as the originals are kept, which have no retention schedule yet. If she declines, the residual risk, losing the originals in a Region-wide loss, is recorded in writing under REL13.

## Decided today

- **Moving the moderation pipeline's Lambda chain onto Step Functions: deferred** by Dev Raman. A multi-quarter rewrite of the pipeline that the Trust and Safety team depends on every day. Left as a medium risk (RELO4).
- **Multi-Region active-active: declined** by Dev Raman. The four-hour recovery time is for failures inside us-east-1 and is planned as a restore. For a Region loss Tidepost accepts a restore in us-west-2 from the daily copies, and active-active would double the estate.

## What the Remediation adds to the bill

At list prices, the designs agreed today add \$10,300 a month and remove \$790, and production's PostgreSQL 13 Extended Support, \$2,336 a month, stops at the cutover. Staging's \$584 continues, so the bill rises by a net \$10,094 a month, almost all of it the second and third zones.

Each change priced at AWS list prices for us-east-1, with the quantities in its design. Production's Extended Support (\$2,336 a month) is avoided, not removed: it began on March 1 and is in no earlier bill.

| Change                                 | Kind  | A month at list |
|----------------------------------------|-------|-----------------|
| ElastiCache replicas                   | Added | \$2,397         |
| Three more EKS nodes                   | Added | \$1,836         |
| Three more API instances               | Added | \$1,551         |
| Cross-zone traffic                     | Added | \$1,536         |
| Production PostgreSQL 16 runs Multi-AZ | Added | \$1,313         |
| AWS Backup                             | Added | \$512           |
| DynamoDB point-in-time recovery        | Added | \$410           |
| OpenSearch dedicated masters           | Added | \$280           |

| Change                                                     | Kind      | A month at list |
|------------------------------------------------------------|-----------|-----------------|
| Allocated database storage doubled                         | Added     | \$236           |
| Cross-Region backup copies                                 | Added     | \$164           |
| NAT gateways in three zones                                | Added     | \$66            |
| Production EKS out of extended support                     | Removed   | -\$365          |
| Ranking endpoint moved to CPU                              | Removed   | -\$425          |
| PostgreSQL 13 Extended Support on staging, which continues | Continues | \$584           |
| Net change                                                 |           | \$10,094        |

Gwen Tallis will see the increase in the April bill. The cost engagement starts from that bill, so none of it is hidden from the savings it counts.

## The four weeks

The sequence agreed today. Reliability's network and backups come first, because the zone moves and the cutover depend on them.

| WeekWhat closes                                                   | Acceptances                          |
|-------------------------------------------------------------------|--------------------------------------|
| 3RELO1, RELO2, COST02, COST03                                     | PERF03, COST06, COST07, SUS02, SUS04 |
| 4OPS04, OPS07, SEC07, SEC11, RELO7, REL11, PERF02, COST01, COST04 | None                                 |
| 5OPS05, OPS06, OPS10, RELO9, REL10, REL12, REL13                  | None                                 |
| 6Verification of every fix, milestone two, the handover           | None                                 |

Kofi Brandvold owns the sequence day to day. The weekly check-in with you is thirty minutes, and the change reviews run twice a week through week five.

# The gate and the fee

---

**Every agreed HRI is closed or accepted in writing, verified in milestone two.** Milestone two is saved in week six, on March 30, 2026. The fee is \$29,750 (\$35,000 at list, less 15%): \$14,875 was invoiced at signature, and \$14,875 is invoiced only when the gate is met. The 21 MRIs are not part of the gate; they stay in the Tool's improvement plan, and the review report lists each one.

---

© 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.