Vantalect

AWS COST · SAMPLE DELIVERABLE

Analysis Sponsor Readout

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

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

September 4, 2026 · AWS Cost Analysis & Optimization: The Optimization (six weeks)

SAMPLE DELIVERABLE

This is a sample deliverable. StrataForge Robotics, Inc. is a fictional company: every account, resource, finding, figure, and person in this report is simulated to show the form and depth of a Vantalect engagement. AWS prices are list prices from the AWS Price List API on October 13, 2026.

Contents

The number 2

How much to plan on 5

The top 10 5

Decisions we need from you 6

The risks 8

The private pricing issue 9

What the next four weeks put in place 11

The guarantee gate 12

The number

The Analysis identifies $23.5 million a year of savings. Plan on $17.2 million to $21.6 million of it, and know that at today's usage only $13.7 million is realizable until StrataForge renegotiates its private pricing commitment. Priya Raman and Ravi Menon: this is the two-week Analysis in one document. It finds $23.5 million a year in 13,020 opportunities on 73,592 resources. If every one lands, the AWS bill falls from $4.6 million a month to $2.6 million, a reduction of 43.5%, because Support falls with the bill too. The range is a planning assumption about how much of each kind of change reaches the bill, and the realizable figure is the private pricing floor at work. Both are estimates from your usage and AWS list prices, net of the 7% discount. Week 6 turns them into a measurement on your Cost and Usage Report.

The bill it starts from is the one the board asked about: $4.6 million in July 2026, 49% above August 2025, against 9% revenue growth. Left alone, it reaches $7.1 million a month by August 2027.

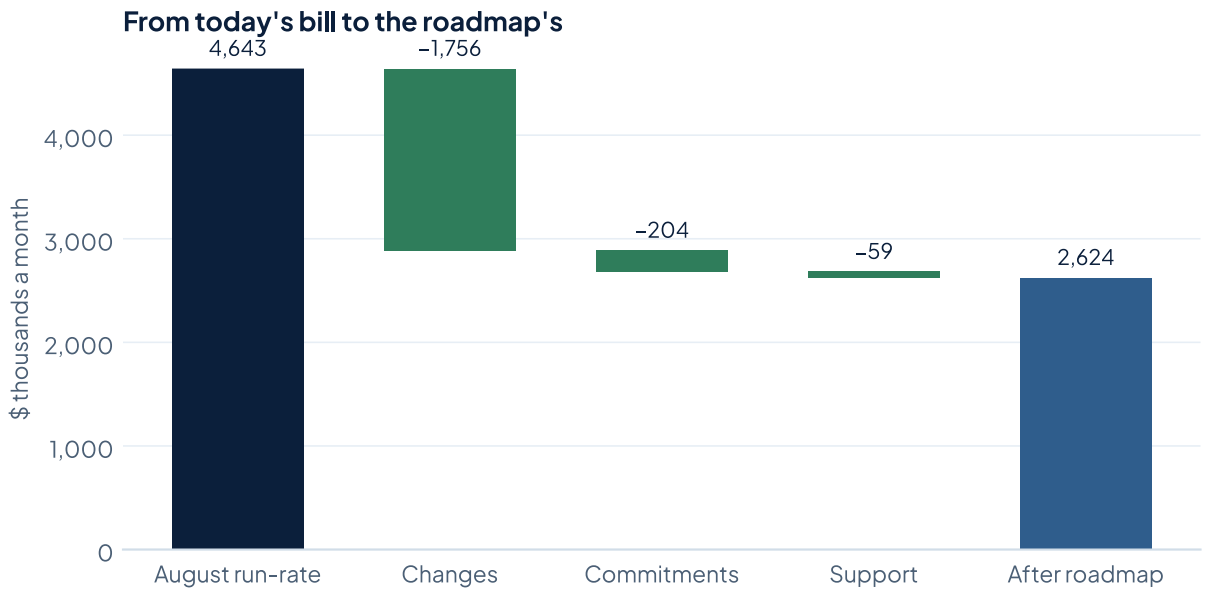
The number, net of the private pricing discount. Identified, range, and realizable are three different questions.

Measure	Per month	Per year
August 2026 run-rate	$4,642,925	$55,715,103
After the roadmap	$2,623,719	$31,484,631
Reduction in the bill, Support included	**$2,019,206**	**$24,230,472**
Identified savings (changes and commitments)	$1,960,018	$23,520,211
Planning range, low to high		$17,239,885 to $21,617,305
Realizable at today's usage, private pricing unchanged		$13,715,103
Realizable at today's usage, with the commitment reset (next sections)		$23,791,047
One-time cost to get there		$128,812

Why the realizable figure with the commitment reset is higher than the identified figure. The two measure different things. Identified is the sum of the changes, priced at the August run-rate. Realizable is what the bill falls by in contract year 3 (February 2027 through January 2028) at today's usage, projected on the roadmap as planned. That year-3 reduction also includes the Support fee that falls with the bill, and it counts the backlog changes only for the part of the year after they land. Nothing in it assumes a change is declined: a readout in week 2 cannot know which changes owners will decline, and week 6 reports those.

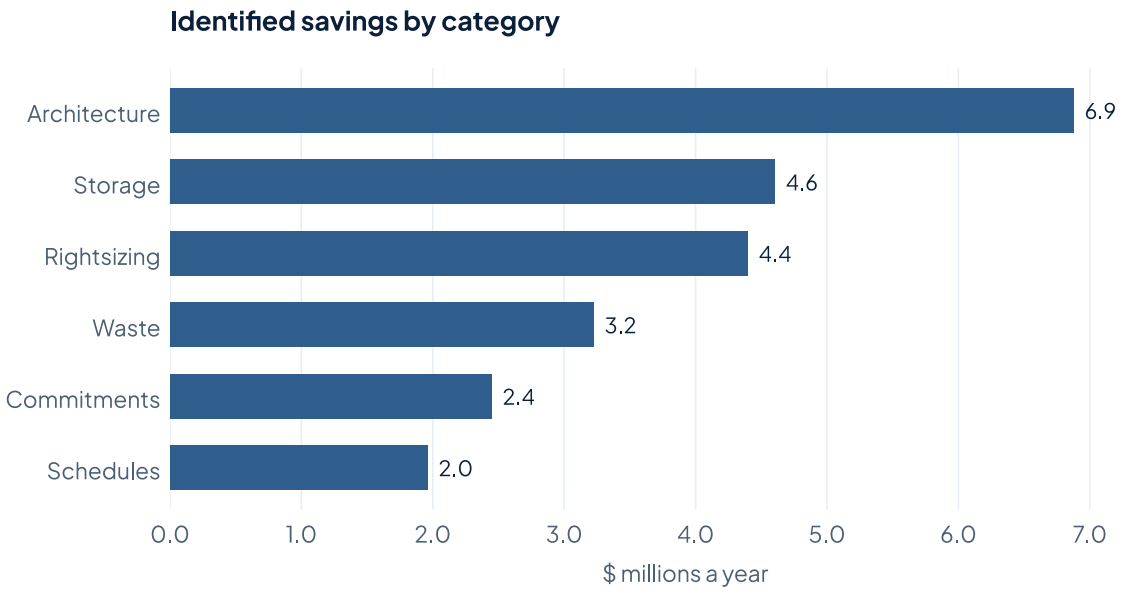
From the identified savings to the realizable figures. With the commitment reset there is no shortfall, so the realizable figure with the reset equals the year-3 reduction.

From identified to realizable, per year	Amount
Identified savings, changes and commitments	$23,520,211
Add: Support falls with the bill	$705,742
Add: the coverage effect, because the changes overlap slightly	$4,519
Less: backlog changes that start partway through contract year 3	$439,425
Reduction in contract-year-3 spend at today's usage	**$23,791,047**
Less: shortfall payment billed under the unchanged commitment	$10,075,944
Realizable at today's usage, private pricing unchanged	**$13,715,103**



From today's bill to the roadmap's, in thousands of dollars a month.

The savings come from six places, and 22% of them are waste and non-production schedules that need no design:



Identified savings by category, millions of dollars a year.

How much to plan on

Identified is the full value of every change if it lands as simulated. Not every change does, so we sort the roadmap into four tiers by how much stands between the plan and the bill, and apply a realization assumption to each. The assumptions are ours, set for planning and not measured; week 6 replaces them with measurements.

The roadmap in four tiers, with the planning assumption for each. The assumptions are estimates, not results.

Tier	Changes	Identified perShare that yearreaches the bill	Planning range per year
Hygiene: scripted, low risk, owner sign-off only	10,071	$9,957,05090% to 100%	$8.96 million to $9.96 million
Engineering: needs build, test, and a change window	2,879	$9,540,56060% to 85%	$5.72 million to $8.11 million
Decision or dependency: needs an owner decision, firmware, or a vendor	67	$1,572,74130% to 70%	$471,822 to $1.10 million
Commitments: finance approval, sized to what remains	3	$2,449,86185% to 100%	$2.08 million to $2.45 million
Total	**13,020**	**$23,520,211**	**$17.24 million to $21.62 million**

Why a capable team missed it. Three things outran a capable FinOps team: the Kestrel acquisition added five accounts with never-expire defaults in November 2025, GPU training spend tripled in a year, and platform defaults (no log retention, no lifecycle rules, gp2 launch templates) created waste faster than quarterly reviews could find it.

Read the tiers as the order of work. The first tier is scripted and needs only each owner's yes. The second needs engineers to build, test, and ship it. The third waits on a decision, a firmware release, or a vendor. The fourth is finance's: it is sized last, to the bill that remains.

The top 10

The ten largest opportunities are **$5.1 million a year, 22% of the total.** The rest is a long tail: most of the savings come from one simple change repeated across thousands of resources. The full ranked list, with owners, is in the savings roadmap.

The ten largest opportunities by annual savings.

Rank	Opportunity	Owner	Wave	Risk	Per year
1	Resize the commitments as they renew	Elena Vasquez	Week 4	High	$2,100,001
2	Scope CloudTrail data events with advanced event selectors	Elena Vasquez	Week 5	Low	$599,428
3	Reduce the On-Demand Capacity Reservation to what training uses	Hiro Tanaka	Week 3	Medium	$448,400
4	Stop publishing one CloudWatch metric per robot (fleet-ingest-prod, us-east-1)	Sam Okafor	Week 5	Medium	$413,688
5	Add ECR lifecycle policies	Elena Vasquez	Week 3	Low	$339,005
6	Lower the FSx for Lustre throughput tier to what training reads	Hiro Tanaka	Week 5	Medium	$311,980
7	Publish telemetry through IoT Basic Ingest	Sam Okafor	Backlog	Medium	$273,124
8	Stop publishing one CloudWatch metric per robot (fleet-ingest-prod, eu-central-1)	Sam Okafor	Week 5	Medium	$216,619
9	Buy a Database Savings Plan after the Graviton moves	Elena Vasquez	Week 4	Medium	$215,890
10	Snapshot and delete the unused dc2 Redshift cluster	Lucas Ferreira	Week 3	Low	$156,419

In plain terms: the commitment lines come from sizing plans to the estate that will exist after the changes. The rest are logging and metrics that nobody reads, a GPU reservation larger than the work it serves, a container registry that keeps every build, a throughput tier bought for a peak that happens once, and a cluster nobody has queried in more than a year. None needs StrataForge to slow a release, and the ones that touch production wait for the owner's maintenance window.

Decisions we need from you

Five decisions come out of the Analysis. Two have dates, and the cost of waiting is real.

The decisions, in the order we would take them.

No.	Decision	Who decides, and when
1	Approve the roadmap and its four waves, with each change still approved by the owner of its workload.	Priya Raman, today
2	Approve the Database and SageMaker Savings Plans for week four, and the queued Compute Savings Plan of $530 an hour for November 1.	Ravi Menon, before the week-four commitment decision on September 14, 2026
3	Approve Deep Archive for the disaster-recovery copy of raw telemetry: any object comes back within 48 hours, but a whole-copy restore of 625 million objects costs about $105,543 (bulk retrieval, a week of temporary copies, and the transfer back to the primary Region) and takes about 6 days.	Priya Raman, today
4	Send the data-lake retention changes to the data-governance council (meeting October 14), and SAP backup retention to the SOX controls owner.	Lucas Ferreira, for the council meeting
5	Start the private pricing renegotiation now: at today's usage the roadmap takes contract-year-three spend $10,075,944 under the $42 million commitment.	Ravi Menon and Priya Raman, this week

Two of these deserve a paragraph each.

The commitment decision is the one with a deadline. The one-year Compute Savings Plan at $380 an hour ends October 31, 2026. We recommend a one-year Database Savings Plan at $25 an hour and a one-year SageMaker Savings Plan at $61 an hour in week four, and a queued three-year Compute Savings Plan of $530 an hour starting November 1, 2026, with a second tranche of $210 an hour behind it. Against buying nothing new and holding today's commitments as they are, that is $204,155 a month, $2.4 million a year; against renewing every existing commitment as it is and adding the two new plans, it is $175,000 a month. Be clear about what you are signing. The three-year Compute plans are a $19.4 million obligation over their term. The one-year alternative that covers the same hours commits $8.6 million and costs $176,549 a month more. We tested the size too. Buying 90% of it costs $19,518 a month more than the recommendation and buying 80% costs $41,256 more, because the hours left uncovered pay On-Demand rates; utilization stays at 100% either way. If On-Demand Linux hours fall 20%, the three-year plan runs at 83% utilization, and that is the risk you accept. The sizing counts only Linux EC2, Aurora, RDS, ElastiCache for Valkey, OpenSearch, and SageMaker endpoints. Windows, RHEL, and SUSE EC2, Lambda, Fargate, and DynamoDB are eligible too and were left out, so the plans err small, and the second tranche is the place to extend them. Ravi Menon approves the purchases and your team buys them. Vantalect never does.

The Deep Archive decision is about restore time. A disaster-recovery copy of raw telemetry in Deep Archive costs a fraction of Standard, and any one object comes back within 48 hours. Restoring the whole copy is a different event: 625 million objects, 3.12 PB across eu-west-1 and us-west-2. Submitted with S3 Batch Operations at up to 1,000 requests a second in each Region, it takes about 4.4 days just to submit them and about 6 days in all, and costs about $105,543 in total: $24,271 for Bulk retrieval, $15,937 for the temporary Standard copies that the restored objects occupy for seven days, and $65,335 to transfer the data back to the primary Region. Standard retrieval returns objects within 12 hours and costs $131,701 for the retrieval itself, or $212,973 with the same copies and transfer. If a six-day whole-copy restore is too slow for any data class, tell us now and that class stays in a faster tier.

The risks

We rate every change Low, Medium, or High by how hard it is to reverse. 11,171 are Low, 1,848 are Medium, and 1 is High. The High one is the three-year Compute Savings Plan (C-0002), because once it starts it cannot be returned or changed. Medium covers every change that cannot be undone in minutes: the Database and SageMaker Savings Plans, the S3 rules that permanently expire objects or noncurrent versions, and removing the per-robot metrics that dashboards and alarms read. The risks worth a sponsor's attention are these.

The risks, and what reduces each.

Risk	Why it matters	What we do about it
The private pricing agreement	At today's usage the roadmap takes spend $10.1 million below the $42.0 million commitment in contract year 3, and only $13.7 million of the savings is realizable	Open the renegotiation now (next section)
The commitment calendar	The one-year plan ends October 31, 2026; a late decision leaves usage uncovered	Queue the purchase in advance; it can be deleted until it starts
Commitment size (the one High rating)	The three-year plans are a $19.4 million obligation that cannot be returned, and a drop in usage lowers their utilization	Size them to the post-roadmap floor and review utilization monthly
Production changes	Rightsizing, version upgrades, and data-path changes touch live systems	Non-production first, a 14–day soak, one canary, a rollback written down before the change, and production moves only in the owner's window
Capacity you cannot get back	The p5 reservation shrinks, and released GPU capacity may not return on request	Confirm the training calendar with Hiro Tanaka first, and buy Capacity Blocks for peaks

Risk	Why it matters	What we do about it
Version deadlines	Kubernetes 1.31 clusters are upgraded by AWS at the end of extended support on November 26, 2026, with no rollback	Plan those upgrades with their owners in weeks 3 to 5; production windows before that date are scheduled in the backlog
Retention changes	Lifecycle rules on the data lakes change how long data stays in fast storage	Take them to the data-governance council before anyone approves them
Estimates	These are not measurements	Week 6 measures each change on your Cost and Usage Report

The private pricing issue

The agreement decides how much of the savings StrataForge keeps. It gives a 7% discount for an annual spend commitment of $36.0 million in contract year 1, $42.0 million in year 2, and $42.0 million in year 3, and runs through January 31, 2028. Counted toward the commitment, as we model it: spend net of the private pricing discount, Savings Plans and reservation fees included. Year 1 was met and year 2 is projected to be. Year 3 begins February 1, 2027, and what the roadmap does to it depends on whether spend keeps growing.

The commitment against spend in each contract year. Year 3 is projected at today's usage, with the roadmap in place.

ContractPeriod year		Commitment	SpendMet	Basis
1	February 2025 to January 2026	$36,000,000	$36,449,975	Yes estimated: the CUR export starts in August 2025
2	February 2026 to January 2027	$42,000,000	$44,976,749	Yes six months actual, six projected with the roadmap
3	February 2027 to January 2028	$42,000,000	$31,924,056	No projected at today's usage

Contract year 3 under two usage paths. Realizable savings are the reduction in year-3 spend less any shortfall billed.

Scenario	Monthly growth	Year 3 without the roadmap	Year 3 with the roadmap	Shortfall	Realizable savings
Usage stays at today's level	0.0%	$55,715,103	$31,924,056	$10,075,944	$13,715,103
Usage keeps growing at the last six months' pace	3.6%	$84,307,253	$48,195,008	$0	$36,112,245

At today's usage the roadmap cuts year-3 spend by $23.8 million, and $10.1 million of that comes straight back as a shortfall payment, so $13.7 million is realizable. If spend keeps growing at 3.6% a month, the roadmap still clears the commitment and the full reduction reaches the bill. Nobody can promise which path StrataForge is on, so the sensible move is to stop depending on it: with the year-3 commitment reset, the roadmap is worth $23.8 million at today's usage and the shortfall goes away.

We recommend that Ravi Menon asks AWS to reset the year-3 commitment to the flat projection rounded down to the million, which is $31.0 million. AWS does not always grant that on its own terms, so the table shows what the reset is worth at three discount levels, in case the price of the amendment is a lower discount:

Realizable savings if the year-3 commitment is reset to the flat projection, at three discount levels.

Discount	Year-3 commitment	Year-3 spend	Realizable savings	Over doing nothing
7%	$31,000,000	$31,924,056	$23,791,047	$10,075,944
6%	$31,000,000	$32,254,851	$23,460,252	$9,745,149
5%	$31,000,000	$32,585,645	$23,129,457	$9,414,355

Even at the lowest discount in the table, 5%, the reset is worth $9.4 million a year more than leaving the agreement alone. Ask in this order: reset the year-3 commitment; confirm what counts toward it (Savings Plan and reservation fees, and Support, in particular); and learn whether year-2 spend above the commitment ($3.0 million projected) carries forward. Verify the answer against the account team's eligible-spend report each month and your Cost and Usage Report. We worked from a summary and not the contract, so Ravi Menon's team and counsel have to confirm how the commitment is measured and how a shortfall is billed.

Plan for the three answers AWS can give. A plain reset of the floor, at the same discount, is the best case, and AWS rarely grants it for nothing. More often it trades a lower floor for something it wants: a longer term, or a new agreement signed early.

The three answers, from the best case to AWS declining. The middle row has no dollar figure because it depends on terms that do not exist yet.

If AWS...	What it means for StrataForge	Year-3 commitment	Realizable savings per year
Resets the year-3 commitment, at the same discount	The roadmap reaches the bill in full	$31,000,000	$23,791,047
Renews early, with a longer term	A lower floor in return for more years under the agreement: the total StrataForge commits to over the term grows, so finance weighs that against the savings it protects	Set in the negotiation	Between the two figures on either side, depending on the floor
Declines	The agreement stands as written, and the shortfall is billed	$42,000,000	$13,715,103

If AWS declines, three things still help, and none needs AWS's agreement to start. First, find out how the agreement treats AWS Marketplace. Agreements like this one commonly count purchases of third-party software through AWS Marketplace toward the commitment, up to a cap. StrataForge's software that it buys outside AWS today could move there if the vendor lists it, and the cap sets how much of the gap it can fill. We have not sized this, because it depends on the vendors and the cap in the contract. Second, time any prepaid commitment on purpose. The Savings Plans we recommend are No Upfront, so they pay nothing at purchase and move no spend from one contract year to another; buying more of them does not close the shortfall. A Partial or All Upfront plan pays cash when it starts. Whether that counts when paid or as the plan is used depends on the contract, we have not modeled it, and we do not recommend prepaying only to avoid a shortfall, because it spends cash to avoid a charge. Third, let growth do the work: the shortfall disappears if spend keeps growing at 3.6% a month, so the monthly showback and the year-3 forecast decide how much of this you need.

What the next four weeks put in place

If you approve the roadmap, the Optimization carries it out through StrataForge's own change process, with the engineers who own each workload. Weeks 3 to 5 each end with a written report of what changed and what did not; week 6 measures all of it. Changes to production systems wait for their owners' maintenance windows, so part of the roadmap lands after week 6; that part is the backlog, and your team runs it from the handover playbook.

The Optimization, week by week, with the savings each wave carries.

Week	Starts	What goes in	Annual savings in this wave
3	September 7, 2026	Waste removed after each owner signs off, non-production schedules, S3 lifecycle policies, ECR and log retention	$9.7 million
4	September 14, 2026	Rightsizing and Graviton, databases and caches, and the commitment decision	$6.8 million
5	September 21, 2026	VPC endpoints, Kafka and service routing, telemetry and metrics, CloudTrail and Config scope, version upgrades, budgets, anomaly alerts, tag policy, and showback	$4.4 million
6	September 28, 2026	Every change measured on the Cost and Usage Report and annualized; the handover workshop and the monthly cost review	
After	After October 2, 2026	The backlog: changes that wait on a date, a maintenance window, or a robot firmware release, run by your team	$2.6 million

You will see Tom Becker, our day-to-day contact, in a weekly 30-minute sponsor check-in, and in change reviews and pairing sessions with each workload's owner. Nothing in the roadmap is applied without that owner's yes, and every change carries its rollback.

The guarantee gate

The guarantee

The second half of the Optimization's fee is invoiced only if the annualized savings in place at the end of week six, measured on StrataForge's own Cost and Usage Report, are at least three times the $35,000 fee: **$105,000 a year.** There is no contingency fee and no share of what you save. If the gate is not met, the second invoice is not issued, and you keep everything delivered.

This roadmap identifies $23.5 million a year, which is 224 times the gate, and the low end of the planning range, $17.2 million, is 164 times. The week-two exit, which ends the engagement here if the roadmap does not reach three times the fee, does not apply. What counts at week six is smaller than the roadmap: only what is in place and visible on the bill. A change you approve today and decline later counts at the savings we estimated for it.

What we ask of you at the readout: approve the roadmap and its waves, decide the five items above, and name the person in Ravi Menon's organization who owns the private pricing conversation with AWS. Everything else is already in motion.

© 2026 Vantalect, LLC. Sample deliverable; StrataForge Robotics is fictional.

AWS, Amazon Web Services, AWS Graviton, and the AWS service names are trademarks of Amazon.com, Inc. or its affiliates. Kubernetes is a registered trademark of The Linux Foundation. GitHub is a trademark of GitHub, Inc. NVIDIA and Isaac Sim are trademarks of NVIDIA Corporation. SAP and SAP S/4HANA are trademarks of SAP SE. Vantalect is not affiliated with or endorsed by any of them. StrataForge Robotics is fictional.