Vantalect

AWS COST · SAMPLE DELIVERABLE

Handover Playbook and Monthly Cost Review

Week 6 of 6: Hand over · prepared for StrataForge Robotics, Inc.

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

October 2, 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

Summary 2
What the identified figure became 3
The remaining roadmap 4
The commitment calendar 11
The private pricing plan 14
KPIs and unit costs 17
The monthly cost review 18
Guardrail ownership 20
Running the playbook without Vantalect 21

Summary

The engagement identified **$23.5 million a year** of savings. By the end of week 6, **$11.1 million a year** was in place, about 20% of the bill and mostly hygiene, storage, non-production schedules, and rightsizing, and $9.4 million a year of it was measured on the Cost and Usage Report. **$11.8 million a year** remains, with an owner for each item and a gating event for each date. This report is the playbook that lets StrataForge finish the work and keep the bill down without Vantalect: the remaining roadmap, the commitment calendar, the private pricing plan, the numbers to watch, the monthly meeting that watches them, and who owns each guardrail.

Four things in it need a decision or an action from someone at StrataForge, in this order.

1. **Confirm the queued Compute Savings Plan** before October 31, 2026, when the one-year plan ends. The three-year plan of $530 an hour starts November 1, 2026. Ravi Menon owns the approval, Elena Vasquez owns the queue (The commitment calendar).
2. **Open the private pricing renegotiation now.** At today's usage the roadmap leaves contract year 3 $9.8 million under its commitment, and only $13.7 million of the reduction is realizable until the commitment is reset (The private pricing plan).
3. **Finish the upgrades before November 26, 2026.** AWS upgrades any Kubernetes 1.31 control plane still running that day, with no rollback. Elena Vasquez owns it (The remaining roadmap).
4. **Hold the first monthly cost review** inside the next month, with the agenda in The monthly cost review, so the guardrails have owners who have met.

Where the roadmap stands at the end of week 6. Estimates for what is still to do; measurements for what is done.

Where things stand	Figure
Roadmap identified in the Analysis, annualized	$23,520,211
In place at week 6 (the estimate for each change)	$11,129,490
Measured on the Cost and Usage Report	$9,377,111
Still to put in place	$11,760,651
Declined by owners	$355,158
Approved, then declined	$274,912
Monthly net bill at the start	$4,642,925
Monthly net bill the roadmap projects	$2,623,719

What the identified figure became

Every dollar of the identified figure is accounted for in one of four places. The reconciliation below adds up to the identified figure, and it is the table to open at the monthly review when someone asks where the savings went.

The identified savings, by where they stand. Still to put in place includes the unapplied part of lines that were only partly applied.

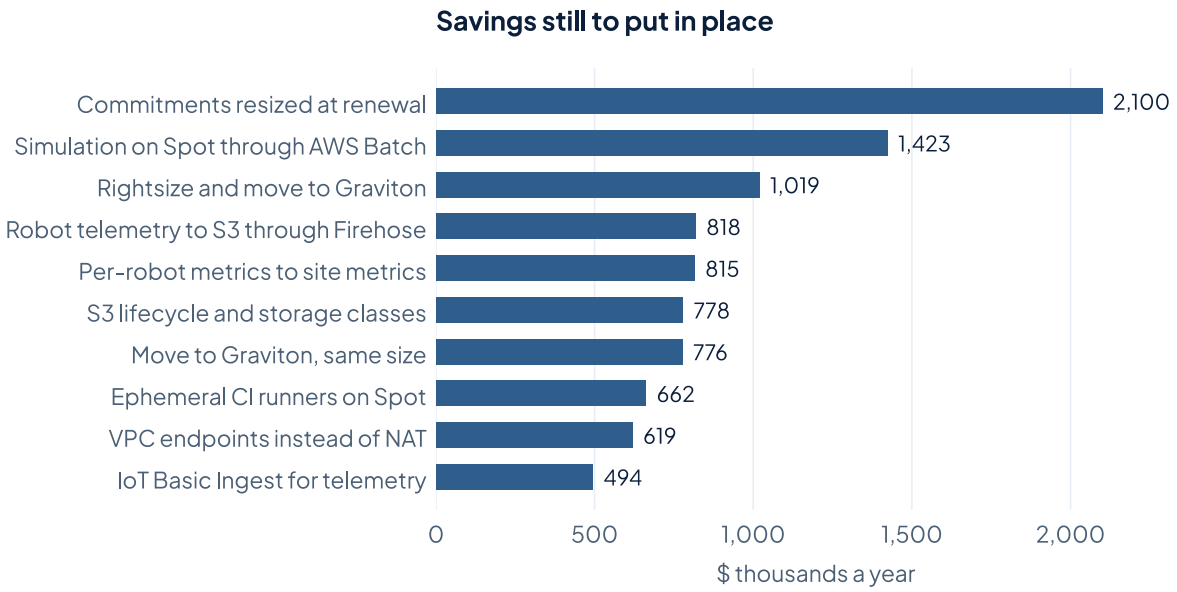
Where it went	Annualized	Share of identified
In place at week 6	$11,129,490	47.3%
Still to put in place	$11,760,651	50.0%
Declined by the owner of the workload	$355,158	1.5%
Approved at the readout, then declined	$274,912	1.2%
Total	**$23,520,211**	**100.0%**

Two notes keep the numbers honest. The in-place figure is an estimate for each change. The week-6 report measured 10,504 of the changes on the Cost and Usage Report at $9.4 million a year, against an estimate of $9.8 million for those same changes. Another $1.3 million is 161 S3 Intelligent-Tiering changes that are in place and not yet visible, so they are not counted. Intelligent-Tiering moves an object only after 30 days without access in the new class (90 for Archive Instant Access), so the first saving shows about October 10 and the full saving by

December 9. Re-measure them with the week-6 method after December 9, 2026. The second note is that a partly applied line, such as the Database Savings Plan, counts its applied part as in place and its unapplied part as remaining.

The remaining roadmap

$11.8 million a year across 28 kinds of change is still to put in place, each waiting on a reason its owner gave. None is abandoned. Production moves wait for their owners' windows, the re-architectures that were merged in week 5 are rolling out through October, and the rest wait on an event, a review, or a build. The gate is stated in each row.



The ten largest groups of savings still to put in place, estimated and annualized.

The remaining roadmap by kind of change. One-time costs are charges that repay themselves within months of the change.

Kind of change	Changes	Annualized	One-time
Commitments resized at renewal	1	$2,100,001	$0
Simulation on Spot through AWS Batch	334	$1,422,523	$0
Rightsize and move to Graviton	290	$1,018,811	$0
Robot telemetry to S3 through Firehose	96	$818,009	$0
Per-robot metrics to site metrics	4	$814,735	$0

Kind of change	Changes	Annualized	One-time
S3 lifecycle and storage classes	53	$778,099	$30,751
Move to Graviton, same size	848	$776,313	$0
Ephemeral CI runners on Spot	143	$661,942	$0
VPC endpoints instead of NAT	47	$619,307	$0
IoT Basic Ingest for telemetry	4	$493,835	$0
Instances off nights and weekends	92	$371,915	$0
FSx for Lustre throughput tier	1	$311,980	$0
Kafka consumers read in their own AZ	4	$194,184	$0
SQL Server Enterprise to Standard	6	$182,814	$0
Database engines out of Extended Support	11	$175,971	$0
Database Savings Plan	1	$164,973	$0
Rightsize (same family)	35	$142,976	$0
ElastiCache: Valkey on Graviton	16	$137,549	$0
Databases off nights and weekends	16	$130,941	$0
Device shadows on change	4	$117,992	$0
Databases: rightsize and Graviton	28	$66,014	$0
All update downloads through CloudFront	4	$63,697	$0
Kubernetes out of extended support	15	$61,101	$0
Lambda on Arm	80	$42,746	$0
OpenSearch: Graviton and gp3	4	$42,204	$0
Service calls stay in their AZ	1	$34,596	$0
MSK brokers on Graviton	4	$15,177	$0
Public IPv4 addresses removed	6	$244	$0
Total	**2,148**	**$11,760,651**	**$30,751**

The same money by what releases it, in date order. The date is the event that releases the item, not a promise on a date we do not control.

The remaining roadmap by the event that releases each part, earliest first.

Target	Waits on	Changes	Annualized	Kinds of change
Oct 6	the gap between training runs on October 6	1	$311,980	FSx for Lustre throughput tier
Oct 8	the change advisory board's October route-table window	47	$619,307	VPC endpoints instead of NAT
Oct 9	a two-week trial in one account	108	$502,856	Instances off nights and weekends; Databases off nights and weekends
Oct 15	the consumers' October release train	4	$194,184	Kafka consumers read in their own AZ
Oct 16	the pilot node group's two weeks in production, then the other node groups	284	$299,064	Move to Graviton, same size; Rightsize and move to Graviton
Oct 17	the October maintenance window	16	$137,549	ElastiCache: Valkey on Graviton
Oct 17	the October maintenance window, readers first	28	$66,014	Databases: rightsize and Graviton
Oct 20	the October OTA release, which moves the download host	4	$63,697	All update downloads through CloudFront
Oct 23	the us-east-1 cutover settling	72	$374,790	Robot telemetry to S3 through Firehose
Oct 23	the owning team's October change window	125	$216,040	Move to Graviton, same size; Rightsize (same family); Rightsize and move to Graviton
Oct 24	a blue/green deployment in the October maintenance window	8	$57,382	OpenSearch: Graviton and gp3; MSK brokers on Graviton
Oct 30	the rollout of the merged change, one Region or service at a time	276	$2,433,328	Ephemeral CI runners on Spot; Simulation on Spot through AWS Batch; Robot telemetry to S3 through Firehose
Oct 30	the us-west-2 Spot pilot's results	234	$959,981	Simulation on Spot through AWS Batch
Oct 30	the data-governance council's review on October 14	51	$658,358	S3 lifecycle and storage classes

Target	Waits on	Changes	AnnualizedKinds of change
Oct 30	the SOX controls owner's sign-off on backup retention	2	$119,741S3 lifecycle and storage classes
Nov 1	the queued purchase's start date	1	$2,100,001Commitments resized at renewal
Nov 10	application testing on the new engine version	11	$175,971Database engines out of Extended Support
Nov 13	an arm64 build of the function's native dependencies	80	$42,746Lambda on Arm
Nov 20	production upgrade windows, before Kubernetes 1.31 is force-upgraded on November 26	11	$44,807Kubernetes out of extended support
Nov 30	robot firmware 7.4, which ships through OTA updates in November	8	$611,827IoT Basic Ingest for telemetry; Device shadows on change
Dec 15	the r5 reservations ending November 30	156	$251,028Move to Graviton, same size; Rightsize and move to Graviton
Dec 18	multi-arch (x86 and Arm) container builds for the service	614	$1,172,213Move to Graviton, same size; Rightsize and move to Graviton
Jan 15	the Aurora reservations ending January 14, 2027, and the Graviton database moves	1	$164,973Database Savings Plan
Mar 31	the MES vendor's certification of SQL Server Standard	6	$182,814SQL Server Enterprise to Standard
	Total	**2,148**	**$11,760,651**

The 30 largest items follow, with who owns each and what it waits on. They are 44% of the remaining value. A line that was only partly applied shows what is in place and counts only the rest here.

The largest items still to put in place, with owner, repository, what each waits on, and the date it is expected.

ID	Change	Owner	Repository	Remaining per year	Waits on	Target
C-0002	Resize the commitments as they renew (sf-management, global)	Elena Vasquez	`platform-landing-zone`	$2,100,001	the queued purchase's start date	Nov 1
A-0157	Stop publishing one CloudWatch metric per robot (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$413,688	the rollout of the merged change, one Region or service at a time	Oct 30
A-0170	Lower the FSx for Lustre throughput tier to what training reads (crucible-training-prod, us-east-1)	Hiro Tanaka	`crucible-infra`	$311,980	the gap between training runs on October 6	Oct 6
A-0162	Publish telemetry through IoT Basic Ingest (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$273,124	robot firmware 7.4, which ships through OTA updates in November	Nov 30
A-0158	Stop publishing one CloudWatch metric per robot (fleet-ingest-prod, eu-central-1)	Sam Okafor	`fleet-infra`	$216,619	the rollout of the merged change, one Region or service at a time	Oct 30
C-0001	Buy a Database Savings Plan after the Graviton moves (sf-management, global) ($50,918 in place)	Elena Vasquez	`platform-landing-zone`	$164,973	the Aurora reservations ending January 14, 2027, and the Graviton database moves	Jan 15
A-0164	Publish telemetry through IoT Basic Ingest (fleet-ingest-prod, eu-central-1)	Sam Okafor	`fleet-infra`	$122,249	robot firmware 7.4, which ships through OTA updates in November	Nov 30
A-0054	Let consumers read from the replica in their own Availability Zone (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$120,394	the consumers' October release train	Oct 15
A-0159	Stop publishing one CloudWatch metric per robot (fleet-ingest-prod, ap-northeast-1)	Sam Okafor	`fleet-infra`	$108,644	the rollout of the merged change, one Region or service at a time	Oct 30
T-2565	Lifecycle policy (crucible-data, us-east-1)	Hiro Tanaka	`crucible-infra`	$103,784	the data-governance council's review on October 14	Oct 30
A-0075	Write robot telemetry to S3 through Firehose, not to CloudWatch Logs (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$99,938	the rollout of the merged change, one Region or service at a time	Oct 30
T-2638	Lifecycle policy (kestrel-data, eu-west-1)	Jonah Reyes	`(not in code)`	$85,053	the data-governance council's review on October 14	Oct 30

ID	Change	Owner	Repository	Remaining per year	Waits on	Target
T-2520	Lifecycle policy (crucible-training-prod, us-east-1)	Hiro Tanaka	`crucible-infra`	$83,277	the data-governance council's review on October 14	Oct 30
A-0160	Stop publishing one CloudWatch metric per robot (fleet-ingest-prod, ap-southeast-1)	Sam Okafor	`fleet-infra`	$75,784	the rollout of the merged change, one Region or service at a time	Oct 30
A-0081	Write robot telemetry to S3 through Firehose, not to CloudWatch Logs (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$74,188	the rollout of the merged change, one Region or service at a time	Oct 30
T-2617	Lifecycle policy (erp-sap-prod, us-east-1)	Ana Kowalski	`enterprise-infra`	$73,430	the SOX controls owner's sign-off on backup retention	Oct 30
A-0094	Write robot telemetry to S3 through Firehose, not to CloudWatch Logs (fleet-ingest-prod, eu-central-1)	Sam Okafor	`fleet-infra`	$68,005	the us-east-1 cutover settling	Oct 23
A-0163	Update device shadows on change, not every minute (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$67,988	robot firmware 7.4, which ships through OTA updates in November	Nov 30
S-0408	Scale the node group to zero outside performance-test windows (fleet-perf, us-east-1)	Sam Okafor	`fleet-infra`	$66,551	a two-week trial in one account	Oct 9
A-0072	Write robot telemetry to S3 through Firehose, not to CloudWatch Logs (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$65,605	the rollout of the merged change, one Region or service at a time	Oct 30
A-0166	Publish telemetry through IoT Basic Ingest (fleet-ingest-prod, ap-northeast-1)	Sam Okafor	`fleet-infra`	$60,599	robot firmware 7.4, which ships through OTA updates in November	Nov 30
A-0070	Write robot telemetry to S3 through Firehose, not to CloudWatch Logs (fleet-ingest-prod, us-east-1)	Sam Okafor	`fleet-infra`	$60,598	the rollout of the merged change, one Region or service at a time	Oct 30
T-2629	Lifecycle policy (kestrel-data, us-east-1)	Jonah Reyes	`(not in code)`	$52,810	the data-governance council's review on October 14	Oct 30
A-0724	Upgrade the database engine out of Extended Support (kestrel-prod, us-east-1)	Jonah Reyes	`(not in code)`	$52,140	application testing on the new engine version	Nov 10
A-0104	Write robot telemetry to S3 through Firehose, not to CloudWatch Logs (fleet-ingest-prod, eu-central-1)	Sam Okafor	`fleet-infra`	$48,806	the us-east-1 cutover settling	Oct 23

ID	Change	Owner	Repository	Remaining per year	Waits on	Target
T-2616	Lifecycle policy (erp-sap-prod, us-east-1)	Ana Kowalski	`enterprise-infra`	$46,311	the SOX controls owner's sign-off on backup retention	Oct 30
S-0395	Scale the node group to zero nights and weekends (fleet-staging, us-east-1)	Sam Okafor	`fleet-infra`	$44,513	a two-week trial in one account	Oct 9
A-0055	Let consumers read from the replica in their own Availability Zone (fleet-ingest-prod, eu-central-1)	Sam Okafor	`fleet-infra`	$42,720	the consumers' October release train	Oct 15
T-2557	Lifecycle policy (crucible-data, us-east-1)	Hiro Tanaka	`crucible-infra`	$38,530	the data-governance council's review on October 14	Oct 30
A-0168	Publish telemetry through IoT Basic Ingest (fleet-ingest-prod, ap-southeast-1)	Sam Okafor	`fleet-infra`	$37,864	robot firmware 7.4, which ships through OTA updates in November	Nov 30

Five patterns recur in the list, and each has a standing answer. The figures are the annualized value and the target dates from the table above.

- **Governance reviews.** 53 changes worth $778,099 wait on the data-governance council or the SOX controls owner, with targets all on October 30, 2026. The lifecycle rules are written and tested in the week-3 pull requests; apply each rule to the bucket the reviewer clears, and verify it with the next S3 Inventory report.
- **Builds for Arm and Graviton.** 1,135 changes worth $1.9 million wait on multi-architecture images, native builds, a pilot node group's first weeks in production, or a reservation's end date, with targets between October 16, 2026 and January 15, 2027. The one production pilot is the fleet-ingest node group of three; the workflow in the week-4 report builds both architectures and publishes one manifest, so adopt it service by service, and each adoption releases that service's items.
- **Maintenance windows and release trains.** 239 changes worth $1.3 million follow the owners' windows, with targets between October 8, 2026 and November 20, 2026. Put each item in the window's change ticket in advance, with its rollback, and run production only in those windows, readers first for databases.
- **Rollouts of merged changes.** 582 changes worth $3.8 million were merged and are rolling out one Region or service at a time, with targets between October 23, 2026 and October 30, 2026. Verify each Region with the daily cost query in the week-6 report before the next one starts.
- **Firmware.** 8 changes worth $611,827, among them the IoT Basic Ingest and device-shadow changes, wait on robot firmware 7.4, which ships by over-the-air update, with targets all on November 30, 2026. They are among the largest items on the list and the least in your control, so treat the firmware date as the critical path.

The remaining 131 changes, worth $3.3 million, wait on a purchase date, a trial, a vendor certification, or an engine test, each named in the table.

The commitment calendar

Commitments end on dates, and each ending is a decision with a deadline. Missing one costs money: a Savings Plan that ends with nothing queued leaves the usage it covered at on-demand rates the next hour. The calendar below is the whole list. Ravi Menon approves every purchase, and Elena Vasquez's team queues and buys it in StrataForge's account; the owner of the workload a reservation covers confirms what runs on it before it is renewed or allowed to end.

Every commitment date and what to do at it. Approval of each purchase rests with Ravi Menon.

Date	Event	Why	Status at handover
September 18, 2026	Buy a one-year Database Savings Plan at $25 an hour	covers the Graviton databases, Valkey caches, and OpenSearch domains in place after week four	Done
September 18, 2026	Buy a one-year SageMaker Savings Plan at $61 an hour	the always-on inference endpoints, which run every hour	Done
October 31, 2026	One-year Compute Savings Plan ($380 an hour) ends	do not renew it as it is	Ahead
November 1, 2026	Queued three-year Compute Savings Plan starts at $530 an hour	sized to the post-roadmap floor, net of the 2024 plan	Ahead
November 26, 2026	EKS upgrades any Kubernetes 1.31 control plane still running	extended support for 1.31 ends; upgrade on your own schedule before then	Ahead
November 30, 2026	Twelve r5.8xlarge reservations end	the EMR fleet moves to Graviton in December	Ahead
January 14, 2027	The Aurora reservations end	move those instances to Graviton generation 7, then buy a second one-year Database Savings Plan of $81 an hour (a plan cannot be changed once bought), for $106 an hour in all	Ahead
February 28, 2027	Three-year Compute Savings Plan from 2024 ($210 an hour) ends	replace with a three-year plan at $210 an hour, queued for March 1	Ahead
May 31, 2027	SAP HANA reservations (2 x u-6tb1.112xlarge) end	renew for three years; the HANA hosts are steady	Ahead

For every date, follow the same five steps, starting 30 days before.

1. **Measure the floor.** Take the lowest sustained hourly usage of the covered service over the last 90 days, after the roadmap changes that have landed. A commitment is sized to the floor, never the average.
2. **Read the recommendation.** Run `aws ce get-savings-plans-purchase-recommendation` for the plan type, and read the same figure in Cost Optimization Hub.

3. **Compare options in a short memo:** renew as it is, resize to the floor, a shorter term, or let it lapse. The week-4 memo is the template. Say what happens to the bill each month under each, and test the size: the three-year Compute plans were checked at two smaller sizes, 90% and 80% of their size, and against a 20% fall in On-Demand Linux hours, where the three-year plan would run at 83%.
4. **Approve and queue.** Ravi Menon approves. The purchase is queued with a purchase time that is the moment the old plan ends, so there is no gap, and it can be deleted until it starts.
5. **Record it** in the commitment register: plan, hourly amount, term, start, end, approver, and the date of the next decision.

Check how well the commitments fit every month with the utilization and coverage reports, which are in the monthly review. The Savings Plans coverage report spans all eligible spend, leaves out reservations, and cannot filter by operating system, so it cannot test the Linux EC2 target; the query after these commands measures that one on the Cost and Usage Report.

```bash
aws ce get-savings-plans-utilization --time-period Start=<first-of-month>,End=<first-of-next-month> --granularity MONTHLY
# Coverage of all eligible spend, not the Linux EC2 target (the query below measures that)
aws ce get-savings-plans-coverage --time-period Start=<first-of-month>,End=<first-of-next-month> --granularity MONTHLY
aws ce get-reservation-utilization --time-period Start=<first-of-month>,End=<first-of-next-month> --granularity MONTHLY
aws savingsplans describe-savings-plans --states queued active
```

```sql
-- Linux On-Demand EC2 pool coverage: usage held by a Savings Plan or reservation, over all Linux instance usage,
-- valued at public On-Demand prices. Spot (SpotUsage) and other operating systems are outside the pool.
SELECT round(sum(CASE WHEN line_item_line_item_type IN ('SavingsPlanCoveredUsage', 'DiscountedUsage')
                      THEN pricing_public_on_demand_cost ELSE 0 END)
             / sum(pricing_public_on_demand_cost), 4) AS linux_ec2_coverage
FROM cur2.strataforge_cur
WHERE billing_period = '<this-yyyy-mm>'
  AND line_item_product_code = 'AmazonEC2'
  AND line_item_usage_type LIKE '%BoxUsage%'
  AND element_at(product, 'operating_system') = 'Linux'
  AND line_item_line_item_type IN ('Usage', 'SavingsPlanCoveredUsage', 'DiscountedUsage');
```

Utilization below 100% means StrataForge is paying for hours it does not use; coverage below the target means it is paying on-demand rates for hours it could have committed. The coverage target is 84.9% of the Linux On-Demand EC2 pool, from 33.9% today, and the three-year plan's utilization should read 100%. A month under that goes on the agenda.

The private pricing plan

StrataForge's private pricing agreement gives a 7% discount on eligible usage in return for an annual commitment of $36.0 million in contract year 1, $42.0 million in year 2, and $42.0 million in year 3, for the term February 1, 2025 through January 31, 2028. The roadmap changes the arithmetic of it. Here is the plain version.

The commitment against spend in each contract year. Year 3 is projected at today's usage with the roadmap in place; projections are estimates and move with growth.

ContractPeriod year	Commitment	SpendMet	Basis
1February 2025 to January 2026	$36,000,000	$36,449,975Yes	estimated: the CUR export starts in August 2025
2February 2026 to January 2027	$42,000,000	$45,675,602Yes	six months actual, six projected with the roadmap
3February 2027 to January 2028	$42,000,000	$32,164,611No	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	$32,164,611	$9,835,389	$13,715,103
Usage keeps growing at the last six months' pace	3.6%	$84,307,253	$48,663,133	$0	$35,644,120

The shortfall

At today's usage, contract year 3 spend falls $9.8 million (23%) short of the commitment, and that much of the $23.6 million reduction comes back as a payment, leaving $13.7 million. If spend keeps growing at 3.6% a month, the commitment is cleared and nothing is owed. The terms of the agreement decide what a shortfall costs; read them with counsel before the numbers harden. Do not buy usage you do not need to close the gap: it would spend a dollar to avoid paying a dollar.

The year-3 reduction ($23.6 million) is not the identified figure ($23.5 million). Support falls with the bill ($705,742 more), the changes overlap slightly ($4,519 more), and the changes still to put in place start partway through the year and the declined ones never save ($679,980 less).

This is the cost of success. Cutting the bill by the amount the roadmap does takes spend below what the commitment assumed, unless usage grows. We recommend four things.

1. **Open the renegotiation now,** not at renewal. Ravi Menon leads it with Priya Raman and the AWS account team, and Tom Becker supplies the forecast from the monthly showback. Ask for the year-3 commitment to be reset to $32.0 million, the flat projection rounded down to the million, at the same discount, in an amendment to the agreement. That is the best case; AWS usually wants something in return.
2. **Know the other two answers.** AWS may trade a lower floor for a longer term, or decline. The table after this list sets the three side by side, and a reset at a lower discount is still worth taking.
3. **Find out what counts.** We model the commitment as spend net of the private pricing discount, Savings Plans and reservation fees included. Which charges count, Savings Plan fees and AWS Marketplace purchases among them, depends on the text of the agreement. Ask the account team for the eligible-spend report by month, and track it beside the Cost and Usage Report.
4. **Ask whether year-2 overspend carries forward.** Year 2 is projected at $45.7 million against $42.0 million, $3.7 million over. If it carries into year 3, the shortfall shrinks by that much.

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

Discount	Year-3 commitment	Year-3 spend	Realizable savings	Over doing nothing
7%	$32,000,000	$32,164,611	$23,550,492	$9,835,389
6%	$32,000,000	$32,497,898	$23,217,204	$9,502,102
5%	$32,000,000	$32,831,186	$22,883,917	$9,168,814

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	$32,000,000	$23,550,492
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 the lower floor 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 levers remain, and none needs AWS's agreement to start. Ask whether the agreement counts purchases of third-party software through AWS Marketplace toward the commitment; agreements like this one commonly do, up to a cap, and software StrataForge buys outside AWS today could move there if the vendor lists it. We have not sized it, because it depends on the vendors and the cap. Time any prepaid commitment on purpose: the Savings Plans on the calendar are No Upfront, pay nothing at purchase, and move no spend between contract years, so buying more of them does not close a 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, and prepaying only to avoid a shortfall spends cash to avoid a charge. Last, let growth close the gap: the shortfall is zero if spend keeps growing at 3.6% a month.

Track progress with the CUR, with the caveat that the CUR does not know which usage is eligible. Reconcile it against the account team's report each month; the report decides.

```sql
SELECT date_format(line_item_usage_start_date, '%Y-%m') AS month,
       round(sum(line_item_net_unblended_cost), 0) AS net_spend
FROM cur2.strataforge_cur
WHERE line_item_usage_start_date >= timestamp '<contract-year-start>'
  AND line_item_line_item_type NOT IN ('Tax', 'Credit', 'Refund')
GROUP BY 1 ORDER BY 1;
```

The Support fee follows the bill too: AWS charges Enterprise Support as a share of monthly charges (3% above $1 million a month), so the roadmap lowers it by about $58,812 a month at its end. It is one of the few places the savings are larger than the measurement shows.

KPIs and unit costs

A bill that falls is only good news if the business grows underneath it. Six numbers tell you whether the bill is healthy, and they are the ones the review reads first. The targets come from the after-state of the roadmap, not from a round number.

The six measures the monthly review reads first, with where each stands and where the roadmap's after-state puts it.

Measure	Now	Target	Basis
Fleet Cloud cost per connected robot per month	$10.18	$5.45	after the full roadmap, with the recommended commitments
Simulation cost per thousand episodes	$7.89	$4.34	down 45% once simulation runs on Spot
Commitment coverage of the Linux on-demand pool (Savings Plans and reservations)	33.9%	84.9%	reviewed monthly
Share of spend with a Team tag	75.7%	95% by December	tag policy enforced
Untagged new resources a week	First reading at the first review	zero	tag policy enforced
Monthly bill, net	$4,642,925	$2,623,719	after the full roadmap

The two unit costs matter most because they separate a bill that grew from a business that grew. **Fleet Cloud cost per connected robot per month** is the net cost of the Fleet Cloud accounts divided by the robots connected that month, and it stood at $10.18 at the start. **Simulation cost per thousand episodes** is the Crucible simulation accounts' cost divided by the episodes run, in thousands. A rising bill with a falling unit cost is growth. A rising bill with a flat or rising unit cost is waste coming back.

The showback query in the week-5 report computes the first. Both need the same two inputs: the team's cost from the CUR 2.0 export, and a denominator from the system that counts the thing (the fleet database for robots, the simulation scheduler for episodes). Put the denominators in a small table, `finops.unit_denominators(month, robots_connected, sim_episodes)`, loaded by a monthly job, so the query never depends on someone remembering a number.

The monthly cost review

One meeting a month, 60 minutes, in the second week, after AWS has finalized the previous month's report. It is how everything in this engagement stays done.

Who attends the monthly cost review and what each brings. Security and SAP Basis join when a change touches their guardrails.

Role	Who	Brings
Chair	Tom Becker, Director, Cloud Financial Management	The agenda, the showback, the register of changes
Finance	Ravi Menon, Vice President, Finance	Commitment approvals, the private pricing position
Platform	Elena Vasquez, Vice President, Cloud Platform	The guardrails, the tag policy, the endpoints
Fleet Cloud	Sam Okafor, Director, Fleet Cloud Engineering	Fleet Cloud accounts, firmware dependencies
Crucible	Hiro Tanaka, Director, Crucible Platform	Simulation and training, Spot, reservations
Data Platform	Lucas Ferreira, Director, Data Platform	The lake, retention, Redshift, EMR
Enterprise	Mei Chen, Director, Enterprise Applications	SAP, PLM, MES, and corporate applications
Perception	Jonah Reyes, Director, Perception Platform (formerly Kestrel Vision)	The acquired accounts still outside code
Sponsor	Priya Raman, Chief Technology Officer	Decisions that cross teams; attends quarterly

The agenda, in order. Each item has one lead and one input, so nobody comes without something to read.

Minutes	Item	Lead	Input
5	The month in numbers: bill, forecast, KPIs	Tom Becker	Cost Explorer monthly view; the six measures
10	Showback by team and what moved	Tom Becker	The showback query; top movers by service
10	Anomalies and budgets that fired	Tom Becker	The anomaly digest and budget alerts, each with its owner's answer
10	Roadmap status	Each owner	The register of changes; the remaining roadmap
10	Commitments	Ravi Menon	Utilization and coverage; the next date on the calendar
10	Waste and recommendations	Elena Vasquez	Cost Optimization Hub and Compute Optimizer findings; the waste register re-run
5	Decisions and owners recorded	Tom Becker	The decision list, read back

The views to bring. Cost Explorer: the monthly view grouped by service and by `Team` tag, the Savings Plans utilization and coverage reports, and the reservation utilization report. Athena on the CUR 2.0 export: the showback, the untagged share by account, and the top movers.

```sql
-- Top movers: services whose usage cost changed most against the previous month.
-- Usage lines only, valued at public On-Demand prices, so tax, credits, and commitment fees do
not move the ranking.
WITH m AS (
  SELECT billing_period, product_servicecode AS service, sum(pricing_public_on_demand_cost) AS
cost
  FROM cur2.strataforge_cur
  WHERE billing_period IN ('<previous-yyyy-mm>', '<this-yyyy-mm>')
    AND line_item_line_item_type IN ('Usage', 'DiscountedUsage', 'SavingsPlanCoveredUsage')
  GROUP BY 1, 2)
SELECT coalesce(cur.service, prev.service) AS service,
       round(coalesce(prev.cost, 0), 0) AS previous, round(coalesce(cur.cost, 0), 0) AS
this_month,
       round(coalesce(cur.cost, 0) - coalesce(prev.cost, 0), 0) AS change
FROM (SELECT * FROM m WHERE billing_period = '<this-yyyy-mm>') cur
FULL OUTER JOIN (SELECT * FROM m WHERE billing_period = '<previous-yyyy-mm>') prev ON
cur.service = prev.service
ORDER BY abs(coalesce(cur.cost, 0) - coalesce(prev.cost, 0)) DESC
LIMIT 15;
```

```sql
-- The untagged share by account: usage lines only, valued at public On-Demand prices.
SELECT line_item_usage_account_id AS account,
       round(sum(CASE WHEN element_at(resource_tags, 'user_team') IS NULL THEN
pricing_public_on_demand_cost ELSE 0 END), 0) AS untagged,
       round(sum(pricing_public_on_demand_cost), 0) AS total
FROM cur2.strataforge_cur
WHERE billing_period = '<this-yyyy-mm>'
  AND line_item_line_item_type IN ('Usage', 'DiscountedUsage', 'SavingsPlanCoveredUsage')
GROUP BY 1
ORDER BY untagged DESC
LIMIT 20;
```

```bash
aws cost-optimization-hub list-recommendation-summaries --group-by ResourceType --region us-
east-1
aws cost-optimization-hub list-recommendations --max-results 25 --region us-east-1
aws compute-optimizer get-ec2-instance-recommendations --account-ids <account-id> --max-results
25 --region <region>
```

The two queries rank usage at public On-Demand prices on purpose: commitment fees, tax, and credits would otherwise swamp the ranking. The showback in the week-5 report is the view that spreads commitment fees across the teams that use them. Cost Optimization Hub answers only in `us-east-1`, and Compute Optimizer answers one Region at a time and, from the management account, only for the member accounts you list in `--account-ids`, so loop over the Regions that hold instances.

The decisions the review makes. Every month it decides four things and writes them in the register: which remaining items to release and who does them, whether an anomaly or budget breach is expected or a defect, whether any guardrail threshold changes (with the reason), and whether a commitment decision is now due. A meeting that ends without an owner and a date on each decision has not made it.

Guardrail ownership

Each guardrail was put in place in week 5, and each has an owner, a signal, and a response. They are the reason the bill stays down.

The guardrails, with owner and the response when each fires.

Guardrail	What it does	Owner	When it fires, do this
AWS Budgets per account and per team	48 account budgets and 24 team budgets on the Team tag, alerting owners at 80% and 100% of forecast	Tom Becker	Ask the owner named on the budget whether the forecast is expected. If it is, raise the budget with a reason in the pull request. If not, treat it as an anomaly.
Cost Anomaly Detection	monitors per service and per Team tag, daily summaries to each owner, immediate alerts above $5,000 of impact	Tom Becker	Open the anomaly, find the resource, and decide within one business day: expected, or stop it. An impact above the alert threshold pages the on-call topic.
Tag policy enforced	Organizations tag policy for Team, Application, Environment, and CostCenter with allowed values, enforced for EC2, EBS, RDS, S3, and Lambda	Elena Vasquez	Fix the tag, or add the value to the policy with a reason. The compliance report lists every violation.
Showback report	monthly per-team report from the CUR 2.0 data export and Athena, with unit costs (per robot, per simulation hour, per training GPU-hour)	Tom Becker	If a team has no row, its tag is missing. Fix the tag, not the report.
Default retention	a scheduled check sets 30-day retention on any new CloudWatch log group that has none	Elena Vasquez	No action: the check sets the retention. Review any group the owner sets longer than the default.
gp3 by default	launch templates and the CDK constructs in all four repositories create gp3 volumes; gp2 and io1 creation alerts the owner	Elena Vasquez	Move the volume to gp3, or record why it needs io2. The alert names the creator.

A guardrail whose alerts nobody answers is worse than none, because it teaches people to ignore the channel. If an alert has no useful response, change the alert, do not leave it.

Running the playbook without Vantalect

Everything StrataForge needs is in the four repositories and in this set of reports. Nothing depends on Vantalect holding an account, a role, or a tool. The access role we used is read-only and is deleted at the close.

Where each part of the playbook lives, how often it runs, and who runs it.

Playbook piece	Where it lives	Cadence	Owner
Waste register re-run (Compute Optimizer, Cost Optimization Hub)	`platform-landing-zone/finops/`	Quarterly	Elena Vasquez
Waste removal runbook	`platform-landing-zone/finops/runbooks/waste-remove.sh`	As the register is approved	Elena Vasquez
Non-production schedule stack	`fleet-infra`, `crucible-infra`, `enterprise-infra`	Add an account, add the stack	Each repository's owner
Lifecycle rules (S3, ECR, Backup, logs)	The repository that defines the resource	When a bucket or repository is created	Each repository's owner
Rightsizing, Graviton, Karpenter	The repository that defines the resource	Monthly, from the review	Each workload's owner
Commitment memo and register	The shared finance folder	At each date on the calendar	Ravi Menon
Budgets, anomaly detection	`platform-landing-zone/finops/budgets/`	When an account is added	Tom Becker
Tag policy and the tag-required control	`platform-landing-zone`	Quarterly review	Elena Vasquez
CUR 2.0 export, Athena tables, showback queries	`platform-landing-zone/finops/athena/`	Monthly	Tom Becker
Savings measurement	The query and formula in the week-6 report	Quarterly, and for each large change	Tom Becker

The cycle. Monthly: the review above. Quarterly: re-run the waste register and the rightsizing plan against 30 days of fresh data, re-measure the largest changes with the week-6 method, and review the tag policy. Annually: renegotiate private pricing before it renews, and review every commitment date for the next year.

When you add an account. Attach it to the organization's tag policy, the budgets, and the anomaly monitors, add the schedule stack if it is non-production, and tag it from the first resource. The StackSets in `platform-landing-zone` do the first three automatically for an account in a governed organizational unit.

When something goes wrong. A budget breach or anomaly with no clear owner goes to the chair of the review. A rollback uses the rollback column in the week-3, 4, and 5 reports, which was written for the person who was not in the room. A change that raised the bill is measured and reverted like any other.

When to ask for help. A bill that moves more than a fifth in a quarter with no explanation, a new service at scale with no cost model, or a commitment decision larger than the ones in the calendar are the three cases we would suggest a second opinion. The rest of the playbook should not need one.

What you keep

The handover workshop, its recording and transcript, and every report, query, runbook, and decision are in your Pavilion workspace, and you take one archive of all of it at the close. The access role is removed from your accounts when you tell us. Vantalect's working files are deleted from our systems within thirty days after the close.

© 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.