

AWS COST · SAMPLE DELIVERABLE

Rightsizing and Commitments

Week 4 of 6: Rightsizing and commitments · prepared for
Copperwick Therapeutics, Inc.

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

March 20, 2026 · AWS Cost Analysis & Optimization: The Optimization (six weeks)

SAMPLE DELIVERABLE

This is a sample deliverable. Copperwick Therapeutics, Inc. is a fictional company: its people, programs, products, accounts, and figures are invented for this sample, and any resemblance to a real company, person, program, or medicine 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 17, 2026; Spot prices are 30-day means for September 14 to October 14, 2026.

Contents

Summary	2
What went in	5
The week, day by day	7
B-1: genomics Batch on Spot	8
Why not c6i.8xlarge on Spot	8
The arithmetic	9
F-1: the genomics Lustre cutover	10
S-3a: the GMP validation environments on a weekday schedule	12
The commitment decision memo	14
The numbers	14
Why \$50.00 and not more	17
What was not bought.....	17
The risk, plainly	18
Approval and purchase.....	18
The private pricing agreement, plainly	19
What stays deferred, and what was declined	20
Honest limits, as of March 20, 2026.....	22
What comes next	23

Summary

In week 4 Copperwick put **\$729,373 a year** of savings in place, in 4 changes: the genomics Batch pipelines moved to Spot (B-1), the genomics Lustre file system replaced by one sized to its measured peak (F-1), a weekday schedule on the GMP validation environments (S-3a), and a one-year Compute Savings Plan (C-1). Added to week 3's \$1,796,936 in 18 changes, that makes 22 changes and **\$2,526,309 a year in place, estimated**: 13.8% of the in-scope bill's \$18,315,196 a year. L-4, approved at the readout and reversed by the CISO, still counts toward the gate at its estimate (\$35,194), so 23 changes and \$2,561,503 a year count through week 4, the figure the readout planned. Every saving in this report is **net of Copperwick's 6% private pricing discount**,

because the discount is part of what the bill falls by; a unit price shown on its way to a saving is at list and says so. Nothing is measured until week 6, so every figure is the Analysis's estimate, annualized.

The GMP campaign freeze began on Monday, March 16, and none of this week's changes touches a system inside it. The freeze covers the production manufacturing systems that make and release the campaign's batches, and a freeze stops every change in its scope, with no exception for cost. B-1 and F-1 run in the genomics research account, C-1 is a purchase in cw-cicd, and S-3a is a schedule on the GMP validation environments, which hold no GxP records and are outside the GMP freeze; the GxP change advisory board approved it on Tuesday, March 17, and it executed in the Thursday release window, March 19 ([The week, day by day](#)).

Spot was priced honestly. On-Demand c6i.8xlarge workers cost \$1.36 an hour, and Spot for that same instance type was only 33.6% below On-Demand in the price table's 30-day window, which would not have justified the interruptions. So the new environments draw from two 32-vCPU pools, so every job fits one instance: c7i.8xlarge (a 30-day mean of \$0.5815 an hour, 57.2% below c6i.8xlarge On-Demand) and, as the second pool, c6i.8xlarge. The estimate prices every Spot hour at the c7i.8xlarge mean plus a 12% allowance for retried work ([B-1: genomics Batch on Spot](#)).

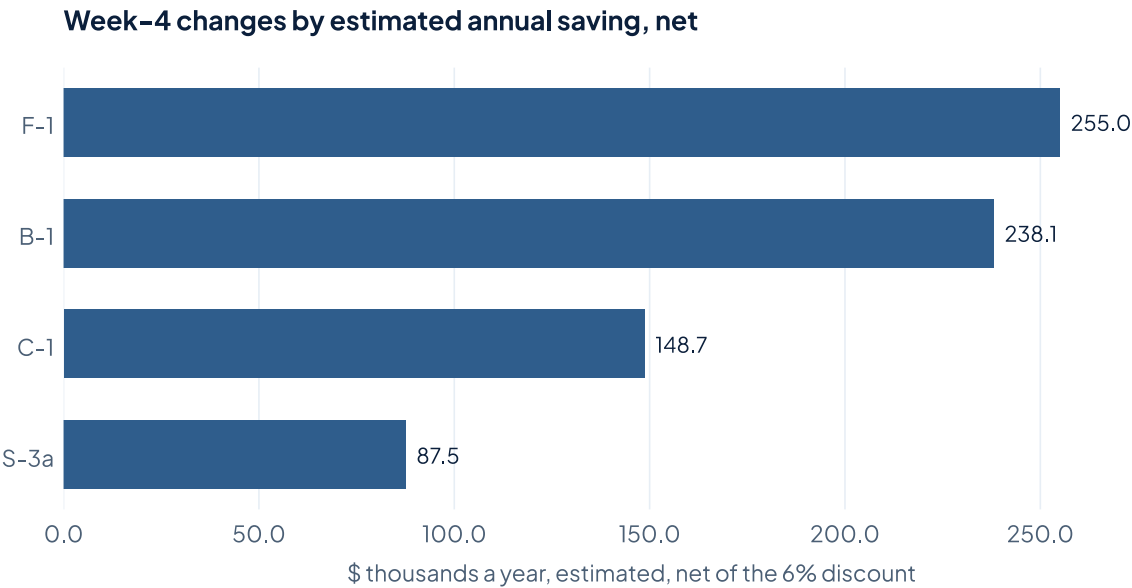
The genomics Lustre file system was provisioned for a peak that never came. It was 64 TiB at 1,000 MB/s/TiB (64 GB/s), and its 90-day throughput peak was 14.8 GB/s. The new file system is 48 TiB at 500 MB/s/TiB (24 GB/s, 62% above the peak). The next step down, 250 MB/s/TiB, would give 12 GB/s, which is 19% below the peak, so it was rejected ([F-1: the genomics Lustre cutover](#)).

The commitment decision was made on Thursday, March 19. Rafael Quintero, Vice President, Financial Planning and Analysis, with Adaeze Whitfield, approved a **one-year No Upfront Compute Savings Plan at \$50.00 an hour** (C-1), bought in cw-cicd that afternoon. The existing three-year plan stays 100% used after every change, a three-year plan waits for the 2027 renewal (in Rafael Quintero's words, "I'm not locking three years a year before I renegotiate the agreement."), and the p4d plan (C-2) was declined at the readout ([The commitment decision memo](#)).

The private pricing agreement does not put the commitment at risk. Copperwick's whole organization spends \$34,964,476 a year net of the discount against a \$27,000,000 commitment. With the full roadmap in place it would still spend \$30,654,432, \$3,654,432 above the commitment, before planned growth ([The private pricing agreement, plainly](#)).

Week 4 in numbers. Savings are estimates, net of the 6% private pricing discount, until the week-6 measurement.

Week 4 at a glance	Result
Changes put in place	4
Annualized savings, estimated, net	\$729,373
Share of the in-scope bill	4.0%
Of which the Savings Plan (C-1), net	\$148,682
In place after weeks 3 and 4, estimated, net	\$2,526,309 (22 changes)
Share of the in-scope bill, all in place	13.8%
Hourly commitment bought	\$50.00 (\$47.00 after the discount)
Existing three-year plan, utilization after every change	100%



The week-4 changes, estimated, annualized, and net of the discount.

Decision for the sponsor

Adaeze Whitfield: no decision is needed to keep this week's savings, and the commitment decision is made and recorded below. Two dates matter next. On Tuesday, March 24, the GxP change advisory board receives the deferred GxP change requests with their validation impact assessments (the week-5 report), and the GMP campaign freeze ends on Friday, March 27. The clinical freeze continues to Friday, April 24, topline.

What went in

The week-4 changes in date order, net of the 6% discount. At list the four are \$775,929 a year; the totals add the whole-dollar rows above them.

ID	Change	Owner	GxP path	Date	Monthly, net	Annualized, net
B-1	Genomics Batch, research pipelines: 40% of their work to Spot (c7i.8xlarge and c6i.8xlarge pools), with a 12% retry allowance	Priya Okonkwo	Non-GxP	Mon, Mar 16	\$19,843	\$238,111
F-1	Genomics Lustre: 64 TiB at 1,000 MB/s/ TiB to 48 TiB at 500 MB/s/TiB (a new file system on the same S3 repository)	Priya Okonkwo	Non-GxP	Tue, Mar 17	\$21,253	\$255,040
C-1	Compute Savings Plan, one year, No Upfront, \$50 an hour, on the non-GPU floor	Rafael Quintero	Non-GxP	Thu, Mar 19	\$12,390	\$148,682
S-3a	GMP val environments (LIMS, MES, CDS): weekday schedule except during executed protocols	Daniel Achterberg	GxP change request	Thu, Mar 19	\$7,295	\$87,540
Total					\$60,781	\$729,373

3 of the 4 are rated High confidence. **B-1** is Medium: its saving depends on Spot prices and the retry rate staying near the figures used ([B-1: genomics Batch on Spot](#)).

What is in place on March 20, 2026, net of the discount. S3-4 is in place but cannot be measured yet, so it is reported separately below and counts for \$0.

Running total, estimated	Changes	Annualized, net	Share of the in-scope bill
In place: week 3	18	\$1,796,936	9.8%
In place: week 4	4	\$729,373	4.0%
In place, counted when measured	22	\$2,526,309	13.8%

At the estimates the 22 changes clear the \$105,000 gate by \$2,421,309. The gate counts only what week 6 measures on Copperwick's Cost and Usage Report, so that margin is a forecast, not a result.

How the roadmap reconciles, net, to 23.5% of the in-scope bill. S3-4's first measurement window runs May 11 to May 17, 2026, with the result on May 19. Deferred and declined changes count for nothing toward the gate.

Where the Analysis's 41 changes stand on Mar 20	Changes	Annualized, net
In place, weeks 3 and 4	22	\$2,526,309
In place, not yet measurable (S3-4): \$0 toward the gate	1	\$119,686
Approved at the readout, deferred, with dates	13	\$1,192,737
Declined at the readout	4	\$436,118
Approved at the readout, then declined (L-4)	1	\$35,194
Identified by the Analysis (36 findings, priced as 41 changes)	41	\$4,310,044

The week, day by day

Week 4, March 16 to 20. The GMP campaign freeze runs from March 16 to March 27.

Date	What happened	Changes
Monday, March 16	The GMP campaign freeze begins; B-1's Spot compute environments go live with per-genome cost tracked daily	B-1
Tuesday, March 17	The change advisory board approves S-3a, to execute in Thursday's release window	S-3a
Tuesday, March 17	F-1: the new genomics Lustre file system takes over between pipeline batches; the old one is deleted only after a full pipeline batch has run on the new one, after a checksum comparison	F-1
Thursday, March 19	S-3a executes in the release window: the weekday schedule starts on the GMP val environments	S-3a
Thursday, March 19	The commitment decision with Rafael Quintero and Adaeze Whitfield: C-1 approved and bought that afternoon in cw-cicd	C-1
Friday, March 20	This report; the GMP campaign freeze continues through Friday, March 27	

What the freeze covers. The GMP campaign freeze, March 16 to March 27, covers the manufacturing systems in production, LIMS, MES, CDS, and the process historian: `gmp-lims-prod`, `gmp-mes-prod`, `gmp-cds-prod`, `gmp-historian-prod`. The clinical freeze, March 2 to April 24, covers the clinical systems from Basalt's database lock to topline: `clin-sce-prod`, `clin-sce-val`, `clin-cdr-prod`, `clin-biomarker-prod`. A freeze covers systems, not families of accounts: development accounts (S-2 scheduled the `clin-` and `gmp-` development accounts, which are in no freeze), the GMP val environments, and clinical operations are outside both. A freeze stops every change in its scope, and cost is not an exception, which is why R-2, S-3b, and E-2b wait for topline ([What stays deferred, and what was declined](#)).

Why S-3a was allowed. The val environments are validation copies of LIMS, MES, and the chromatography data system. They host no GxP records and are not in the campaign's batch path, so they are outside the GMP freeze, and the change is a GxP change request, not a freeze exception: it went to the board with a validation impact assessment, was approved on Tuesday, March 17, and executed in the next release window, Thursday, March 19 ([S-3a: the GMP validation environments on a weekday schedule](#)).

B-1: genomics Batch on Spot

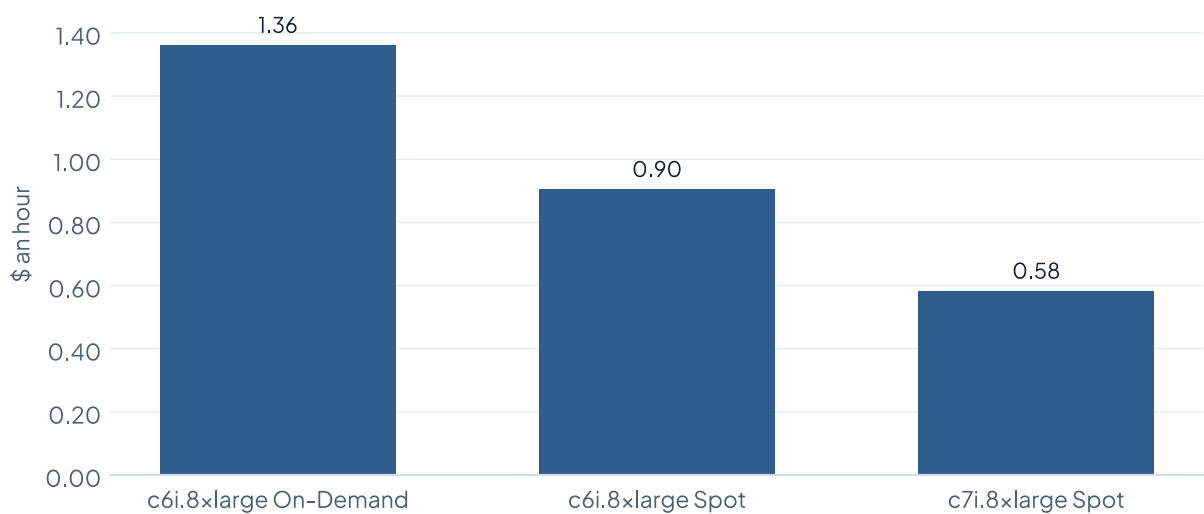
B-1 moved 40% of genomics work from On-Demand to Spot on Monday, March 16. The genomics pipelines run on AWS Batch with 102 On-Demand c6i.8xlarge workers on average. The alignment and variant-calling stages checkpoint and retry under Nextflow, so they tolerate an interruption; that is why those stages move and the rest stays On-Demand. The environments use the price-capacity-optimized allocation strategy across 2 pools, c7i.8xlarge and c6i.8xlarge (32 vCPUs each, so every job fits one instance).

Why not c6i.8xlarge on Spot

Prices at list from the committed price table; Spot is the table’s 30-day mean. The last column compares each to what the workers cost today, \$1.36 an hour.

32 vCPUs of compute	vCPUs and memory	On-Demand an hour	Spot, 30-day mean, an hour	Spot below its own On-Demand	Below c6i.8xlarge On-Demand
c6i.8xlarge, as it runs today	32 and 64 GiB	\$1.360	\$0.9034	33.6%	33.6%
c7i.8xlarge	32 and 64 GiB	\$1.428	\$0.5815	59.3%	57.2%

What 32 vCPUs cost an hour, at list



What 32 vCPUs cost an hour: today’s On-Demand worker against the Spot options.

Spot for c6i.8xlarge itself was only 33.6% below On-Demand, so staying on that type would save too little for the interruptions it brings. Moving to c7i.8xlarge puts the Spot hour 57.2% below today's On-Demand worker. The second pool is the instance type the workers run on today, which keeps capacity available when the c7i pool is short. The estimate prices every Spot hour at the c7i.8xlarge mean, so an hour that lands in the c6i.8xlarge pool (\$0.9034) costs more than the estimate allows, and it gives no credit for c7i's faster cores. Savings Plans do not apply to Spot hours, which is why C-1 was sized after this change.

The arithmetic

B-1 priced from the committed price table. The 12% allowance means every Spot hour is billed as 1.12 hours, to cover work that is retried.

B-1, a month	Figure
Genomics Batch workers, On-Demand c6i.8xlarge, on average	102
Share of the work moved to Spot	40% (41 workers)
Hours a month (41 workers times 730 hours)	29,784
Those hours On-Demand, at \$1.36, at list	\$40,506
Those hours on Spot, at \$0.5815 times 1.12 for retries, at list	\$19,397
Saving a month, at list	\$21,109
Saving a month, net of the 6% discount	\$19,843
Saving a year, net	\$238,111

The result depends on two things the estimate holds still: the Spot price and the retry rate. Each additional 10 points of retries costs \$1,628 a month net (\$1,732 at list), and the saving would vanish only if retries ran past 134%, which is why B-1 is rated Medium and not High: the arithmetic is sound, and the inputs move. Genomics spend also moves with lab throughput, so B-1 is tracked every day from Monday, March 16, as a cost per 30x whole-genome equivalent processed. A slow lab week then does not read as a saving. The week-6 measurement uses the same unit (cost per 30x whole-genome equivalent processed, before and after, times the 90-day average volume).

YAML

```
# rc-genomics-prod: the Spot compute environment (CloudFormation excerpt)
GenomicsSpotEnvironment:
  Type: AWS::Batch::ComputeEnvironment
  Properties:
    Type: MANAGED
    ComputeResources:
      Type: SPOT
      AllocationStrategy: SPOT_PRICE_CAPACITY_OPTIMIZED
      InstanceTypes: [c7i.8xlarge, c6i.8xlarge]
      MinvCpus: 0
      MaxvCpus: 4096
      Subnets: !Ref GenomicsSubnets      # all zones, so every Spot pool is available
      SecurityGroupIds: [!Ref BatchSecurityGroup]
      InstanceRole: !Ref BatchInstanceProfile
```

GROOVY

```
// nextflow.config: the checkpointed stages retry when a Spot host is reclaimed
process {
  withLabel: 'align|call_variants' {
    errorStrategy = 'retry'
    maxRetries    = 3
  }
}
```

How it is undone. The On-Demand environment is still there and still on the job queue's list. To roll back, point the queue's first order at it, or remove the Spot environment from the queue; running jobs finish where they are. Nothing was deleted.

Two choices stay with their owners. B-2, pinning the genomics environments to the Lustre file system's Availability Zone, was declined at the readout by Priya Okonkwo: "One zone means a third of the Spot pools. I'd rather pay the cross-AZ toll than eat interruptions." Pinning to one zone would also cut the number of Spot pools, which is the same reason it conflicts with B-1. B-1b, the same Spot move for the GCP-validated clinical biomarker pipelines (7 of the biomarker pipelines' 18 workers, the same 40% share), waits until May 11 because an instance-type change is a revalidation under their validation plan.

F-1: the genomics Lustre cutover

The genomics file system was 64 TiB at 1,000 MB/s/TiB, which provisions 64 GB/s. CloudWatch shows 29 TiB used and a 90-day throughput peak of 14.8 GB/s: 23% of what was paid for. Lustre storage cannot shrink, so the change is a new persistent-2 file system of 48 TiB at 500 MB/s/TiB, linked to the same S3 prefix, that took over on Tuesday, March 17.

The four persistent-2 throughput tiers at 48 TiB, priced from the committed table at list. A file system that cannot reach the peak slows every pipeline that hits it.

Throughput per TiB, 48 TiB	Total throughput	Against the 14.8 GB/s peak	Monthly, at Decision list
125 MB/s/TiB at \$0.145 a GB-month	6 GB/s	59% below	\$7,127 Rejected: below the measured peak
250 MB/s/TiB at \$0.210 a GB-month	12 GB/s	19% below	\$10,322 Rejected: below the measured peak
500 MB/s/TiB at \$0.340 a GB-month	24 GB/s	62% above	\$16,712 Chosen
1,000 MB/s/TiB at \$0.600 a GB-month	48 GB/s	224% above	\$29,491 Above the peak, costs more than needed

The new file system is 24 GB/s, 62% headroom over the measured peak, and holds the 29 TiB at 60% full. The step to 250 MB/s/TiB would save more and was rejected on the numbers: 12 GB/s sits 19% below the peak, so a pipeline run at the peak would throttle.

F-1 priced as capacity times price per GB-month, at 64 and 48 TiB of 1,024 GB each. Its one-time charge is the repository import requests, \$40, listed beside the result and never annualized.

F-1, a month	Figure
Old: 64 TiB at \$0.60 a GB-month, at list	\$39,322
New: 48 TiB at \$0.34 a GB-month, at list	\$16,712
Saving a month, at list	\$22,610
Saving a month, net of the 6% discount	\$21,253
Saving a year, net	\$255,040

How the cutover ran. The cutover was between pipeline batches on Tuesday, so no job was mid-run.

1. The new file system was created at the start of the day and linked to the same S3 prefix, with the existing data repository's metadata imported.
2. The Batch job definitions were pointed at the new mount while the queue was drained.
3. A checksum manifest of the working set on the old file system was compared against the same files read through the new one; every file matched.

4. The old file system stays until a full pipeline batch has run on the new one, and is deleted after that. The day of the change is excluded from both measurement windows, so the days both file systems existed fall in neither.

YAML

```
# rc-genomics-prod: the data repository association on the new file system (excerpt)
GenomicsRepository:
  Type: AWS::FSx::DataRepositoryAssociation
  Properties:
    FileSystemId: !Ref GenomicsLustre
    FileSystemPath: /genomics
    DataRepositoryPath: !Sub s3://${GenomicsBucket}/genomics
    BatchImportMetaDataOnCreate: true
    S3:
      AutoImportPolicy: {Events: [NEW, CHANGED, DELETED]}
      AutoExportPolicy: {Events: [NEW, CHANGED, DELETED]}
```

BASH

```
# on a host that mounts both file systems: compare the working set before deleting the old one
cd /fsx-old && xargs -a /var/tmp/working-set.txt -d '\n' sha256sum > /var/tmp/old.sha256
cd /fsx-new && sha256sum --quiet -c /var/tmp/old.sha256 && echo "every file matches"
```

How it is undone. The rollback is to create a larger file system from the same S3 repository, since Lustre cannot be resized downward or in place. Because the repository holds the data, nothing is stranded. The cryo-EM file system gets a different answer, F-3, a scratch file system for each collection campaign at \$0.14 a GB-month, scheduled with B-3 for April 6. A throughput and capacity review before any new file system is built in week 5 as GR-6.

S-3a: the GMP validation environments on a weekday schedule

S-3a puts the application tiers of the LIMS, MES, and CDS validation environments on a weekday schedule: 06:00 to 20:00 Eastern, Monday through Friday, which is 300 of 730 hours a month (41%). It is a GxP change request, because a schedule changes when a validated system's environments are available. Daniel Achterberg, Director, Manufacturing and Quality Systems, owns the change and Dr. Lorraine Osei, Senior Director, Quality Assurance, Computerized Systems, approved it on one condition: **no environment stops during an executed protocol**. The scheduler enforces that with a hold tag.

S-3a priced as the val tiers' monthly cost times the 430 stopped hours of 730.

S-3a, a month	Figure
The val application tiers, 730 hours, at list	\$13,175
The same tiers, 300 hours, at list	\$5,414
Saving a month, at list	\$7,761
Saving a month, net of the 6% discount	\$7,295
Saving a year, net	\$87,540

The validation impact assessment. The val environments host no GxP records, no production data, and no audit trail the inspector relies on, so no GxP function changes. The risk is operational: a stopped environment during an executed operational qualification protocol would invalidate the run. The board accepted unscripted testing on that basis: the environment owner watched the first stop and start, and checked the first-morning start against the application's health check. The hold tag is the control for the one real risk.

How the hold works. Each val instance carries the tag `Schedule: gmp-val-weekday`. When QA executes a protocol, the validation lead changes the tag to `Schedule: gmp-val-hold` for the protocol's duration. The hold schedule is always running, so the scheduler leaves the environment on until the tag is set back.

YAML

```
# gmp-lims-val: two schedules in Instance Scheduler on AWS (CloudFormation excerpt)
WeekdaySchedule:
  Type: Custom::ServiceInstanceSchedule
  Properties:
    ServiceToken: !ImportValue InstanceSchedulerServiceToken
    NoStackPrefix: "True"
    Name: gmp-val-weekday
    Timezone: America/New_York
    Periods:
      - Description: Weekdays, 14 hours
        BeginTime: "06:00"
        EndTime: "19:59"
        WeekDays: Mon-Fri
HoldSchedule:
  Type: Custom::ServiceInstanceSchedule
  Properties:
    ServiceToken: !ImportValue InstanceSchedulerServiceToken
    NoStackPrefix: "True"
    Name: gmp-val-hold
    Description: QA hold during an executed protocol
    OverrideStatus: running
```

How it is undone. Remove the schedule tag from the val instances; they stay as they are. The same schedule applies to the SCE validation environment (S-3b, \$45,819 a year, net) after topline, on April 27, because the clinical freeze covers it.

The commitment decision memo

The decision

To: Rafael Quintero, Vice President, Financial Planning and Analysis. **Copy:** Adaeze Whitfield, Grace Mbeki.

Decided Thursday, March 19: buy a one-year Compute Savings Plan, No Upfront, at **\$50.00 an hour (C-1)**, a commitment of \$438,000 over the year at list (\$411,720 after the 6% discount). Do not buy a three-year plan or pay All Upfront now. Do not buy the p4d plan (C-2). Copperwick’s finance team bought the plan in `cw-cicd`, a member account inside the Savings Plans sharing group, that afternoon.

The numbers

Copperwick has one commitment today: a three-year No Upfront Compute Savings Plan at \$150.00 an hour, in force from August 1, 2024 to July 31, 2027. At its rate it covers \$295.77 an hour of On-Demand usage, and it sits on the usage with the highest discount first. A floor of non-GPU compute runs every hour of every day, and the changes in place have shrunk it. After them, the floor costs **\$412.64 an hour On-Demand**. Three deferred changes (R-2, B-1b, G-1) will shrink it further, to \$395.50. Copperwick restricts Savings Plans sharing to the 48 in-scope accounts (a cost category) (group sharing since January 12), so the floor is the usage in those accounts, and C-1 is measured on the usage it covers there. The payer account cannot belong to a group, so C-1, like the existing plan, was bought in `cw-cicd`, a member account inside it.

The floor is \$412.64 an hour; the table shows three decimals so the rows add to their total exactly. R-2, B-1b, and G-1 are the three deferred changes that move compute hours.

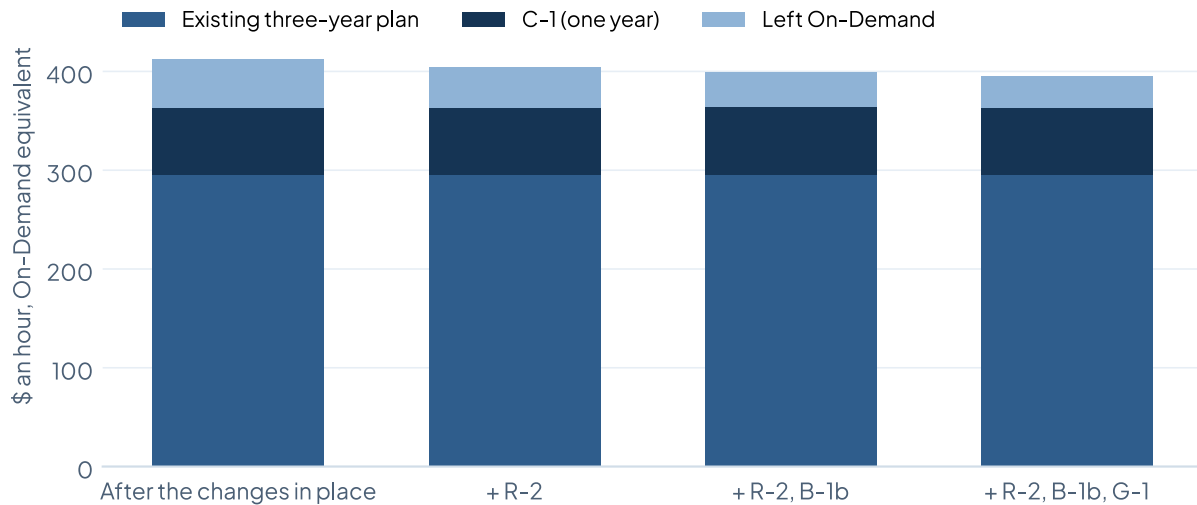
The 24-hour non-GPU On-Demand floor, after the changes in place	An hour, at list
EKS platform On-Demand base (30% of its spend)	\$128.219
Clinical operations apps (80% always on)	\$92.055
Genomics research pipelines On-Demand minimum (60% On-Demand share after B-1, times a 60% minimum)	\$49.939

The 24-hour non-GPU On-Demand floor, after the changes in place	An hour, at list
Clinical biomarker pipelines On-Demand minimum (a 60% minimum)	\$14.688
Validated production (Linux-equivalent compute)	\$41.472
Shared services (90% always on)	\$32.055
Molecular dynamics minimum (4 nodes)	\$21.760
Lab sandboxes minimum (40%)	\$18.630
Data platform services, after R-1	\$13.824
Floor after the changes in place	\$412.642
Less R-2 (SCE grid, 8 to 6 nodes)	-\$8.064
Less B-1b (biomarker pipelines to Spot)	-\$5.875
Less G-1 (molecular dynamics to Graviton)	-\$3.200
Floor after every deferred change	\$395.503

Each deferred change added in turn. An hour is at list. The existing plan stays fully used after every change, and so does C-1.

Estate	Floor, On-Demand	Existing plan covers	C-1 covers	Left On-Both plans Demandused
The estate after the changes in place	\$412.642	\$295.771	\$68.056	\$48.815Yes, 100%
And R-2 (SCE grid, April 27)	\$404.578	\$295.771	\$68.056	\$40.751Yes, 100%
And B-1b (biomarker pipelines to Spot, May 11)	\$398.703	\$295.771	\$68.056	\$34.876Yes, 100%
And G-1 (molecular dynamics to Graviton, April 20)	\$395.503	\$295.771	\$68.056	\$31.676Yes, 100%

The 24-hour floor, covered by the two plans



The 24-hour floor and what the two plans cover, as each deferred change is added.

The existing plan stays 100% used after every change. It covers \$295.77 an hour; the floor after every deferred change is \$395.50. The floor would have to fall by another \$99.73 an hour before the existing plan had an idle hour, so no change's saving is absorbed by an unused commitment. That matters for the week-6 measurement, which counts each change at public On-Demand cost for this reason.

C-1 from the committed price table. A year is 8,760 hours. The saving is the covered On-Demand cost less the fee, times 0.94.

C-1

Figure

Hourly commitment	\$50.00 (\$47.00 after the discount)
Discount at the one-year m6i rate	26.5%
On-Demand usage the commitment covers, an hour	\$68.06
Share of the floor left after every deferred change that C-1 covers	68.2%
What it costs a month, at list	\$36,500
What the covered usage would cost On-Demand a month, at list	\$49,681
Saving a month, net	\$12,390
Saving a year, net	\$148,682

C-1 is a **Compute** Savings Plan, not an EC2 Instance plan, so the Graviton move (G-1) and the Spot move for the biomarker pipelines (B-1b) strand nothing. It was priced at the m6i one-year rate (26.5% off), the lowest of the three families it covers: c6i is 27.7% and r6i 27.6%. Pricing at the lowest rate means the estimate does not depend on which family the plan is applied to first.

The order mattered

The plan was sized after B-1 and R-1, not before. The same floor was \$459.76 an hour before them, \$47.12 an hour more than today's \$412.64, because Spot hours cannot be covered by a Savings Plan and R-1 halved the data platform services. A plan sized on the larger number would have committed to hours those changes removed.

Why \$50.00 and not more

The plan is sized below the floor on purpose. It costs \$50.00 every hour whether or not the usage exists, so it must stay fully used in the worst case, which is after every deferred change lands. Together the two plans cover \$363.83 an hour of the floor's \$395.50, which leaves \$31.68 an hour (8.0%) as a margin. A larger plan would earn more on paper and would idle if any one of the deferred or declined changes grew.

What was not bought

The same \$68.06 an hour of covered usage bought four ways, at the m6i.4xlarge rate in the committed price table. The three-year and All Upfront options save more and commit more.

Option, same covered usage	Outcome	Discount	Commitment, an hour, at list	Saving a month, net	Committed over the term, at list	Paid on day one
One year, No Upfront	Bought (C-1)	26.5%	\$50.00	\$12,390	\$438,000	\$0
One year, All Upfront	Not bought	31.4%	\$46.67	\$14,678	\$408,797	\$408,797
Three years, No Upfront	Waits for the 2027 renewal	49.3%	\$34.51	\$23,016	\$907,043	\$0
Three years, All Upfront	Not bought	54.0%	\$31.32	\$25,209	\$823,066	\$823,066

Three-year plans wait for the 2027 renewal. Rafael Quintero put it this way: "I'm not locking three years a year before I renegotiate the agreement." The private pricing agreement ends June 30, 2027, and a three-year plan bought now would run to March 19, 2029, about 21 months past it. A three-year plan would take 49.3% off the covered usage instead of 26.5%, and waiting costs

about \$10,626 a month net, \$127,513 for the year. That is the price of sizing the commitment and the plan together at the renewal, when the commitment, the discount, and the estate are known. The existing three-year plan ends July 31, 2027, a month after the agreement does.

All Upfront waits too. Paying \$408,797 on day one for a one-year plan would add \$27,451 a year, net. It is a decision about cash, and it belongs to Rafael Quintero.

C-2, the p4d plan, was declined at the readout. The four remaining p4d.24xlarge cost \$87.83 an hour On-Demand; a one-year EC2 Instance Savings Plan for the p4d family would cost \$55.68 an hour and save \$22,059 a month net, \$264,705 a year. Mateo Varga and Rafael Quintero declined it: the training fleet moves to a newer accelerator within the year, and a family-locked plan would strand. The GPU fleet stays On-Demand, and this is a decision, not an oversight. Revisit when the new accelerator is chosen.

Database and SageMaker plans wait. A Database Savings Plan covers Generation 7 and newer database instances only. Every in-scope database runs db.r5 or db.r6i, and R-3 and R-4 keep those generations, so none is eligible; a plan becomes worth sizing only if a database moves to a Generation 7 class. A SageMaker Savings Plan is not worth buying after W-2: the notebooks now bill about 22% of hours.

The risk, plainly

- **A plan can be returned only for 7 days.** AWS allows a return within 7 days of purchase, in the same calendar month, for a plan of \$100 an hour or less; C-1 qualifies, so it could be returned through Thursday, March 26. After that it cannot be canceled, and the \$438,000 (at list) is owed for the year, used or not. The sizing is the only protection.
- **A smaller floor would idle the plan.** The two plans together are fully used while the floor stays above \$363.83 an hour, which is \$31.68 an hour under the lowest the floor is expected to reach. Retiring a workload, or moving one to a service the plan does not cover, would do it.
- **Graviton helps, not hurts.** A Compute plan applies to Graviton as well as x86, so G-1 lowers the bill without stranding it.
- **Support follows usage.** AWS Enterprise Support is a percentage of usage (the top tier is 3%), so every saving lowers it a little. That is not counted.

Approval and purchase

Nothing was bought without Rafael Quintero's approval, and Vantalect buys nothing. Copperwick's finance team made the purchase in cw-cicd, a member account inside the Savings Plans sharing group, which Copperwick restricts to the 48 in-scope accounts (a cost category);

the payer account cannot belong to a group. The steps: read AWS’s own recommendation to compare with ours, find the offering for the term and payment option, buy at the approved hourly rate, and confirm.

```
BASH
aws ce get-savings-plans-purchase-recommendation --savings-plans-type COMPUTE_SP \
  --term-in-years ONE_YEAR --payment-option NO_UPFRONT --lookback-period-in-days SIXTY_DAYS
aws savingsplans describe-savings-plans-offerings --plan-types Compute --durations 31536000 \
  --payment-options "No Upfront" --currencies USD
aws savingsplans create-savings-plan --savings-plan-offering-id <offering-id> \
  --commitment 50.00 --client-token <uuid>
aws savingsplans describe-savings-plans --states active
aws ce get-savings-plans-utilization --time-period Start=<first-day>,End=<last-day> --
granularity DAILY
```

The last command is how utilization is watched: it should read 100% for as long as the floor holds. Record the purchase, the approver, and the date in the commitment register that the monthly cost review keeps.

The private pricing agreement, plainly

Copperwick has a private pricing agreement with AWS: a three-year term from July 1, 2024 to June 30, 2027, a commitment of \$27 million a year in spend net of the discount, and a flat 6% discount on eligible usage, Savings Plan fees included. Every savings figure in these reports is net of that discount, because the discount is part of what the bill falls by.

The question a CFO asks next is whether cutting spend risks a shortfall against the commitment. It does not, and here is the arithmetic. The commitment is measured on the whole organization, not on the 48 accounts in scope: 212 accounts, \$37,196,251 a year at list and **\$34,964,476 a year net of the discount**.

Spend is net of the 6% discount; the savings come off at their net figures. Planned growth adds to every row. No measured result is used: the week-6 report restates the second row on Copperwick’s Cost and Usage Report.

The whole organization, a year, net of the discount	Spend	Above the \$27 million commitment
Today, before any change	\$34,964,476	\$7,964,476
With the 22 changes in place, at their estimates	\$32,438,167	\$5,438,167
With the full roadmap in place (41 changes)	\$30,654,432	\$3,654,432

The entire identified roadmap, \$4,310,044 a year net, is 54% of the \$7,964,476 between today's spend and the commitment. A shortfall would need the organization to spend less than \$27 million in a year, and every row above stays at least \$3,654,432 over it. The savings do inform the 2027 renewal: Copperwick can commit to a lower annual figure, or the same figure at a deeper discount. That value is real and is not counted toward the gate.

What stays deferred, and what was declined

Every deferral has a date and a reason, and counts for nothing toward the gate until it is in place and measured. 13 changes are deferred, worth \$1,192,737 a year net.

The deferral calendar. S-3b, R-2, and E-2b wait for topline on April 24. E-2b covers the clinical systems only; its twin on the manufacturing systems, E-2a, is already in place.

Changes	Date	Waiting on
F-3, B-3	Monday, April 6	The next cryo-EM campaign
N-1b, S3-3	Tuesday, April 14	Approved change requests (S3-3 with a documented retrieval test)
S3-1b	Monday, April 20	The tar job's second pass
G-1	Monday, April 20	The molecular dynamics benchmark
S3-5	Tuesday, April 21	The disaster-recovery test
S-3b, E-2b	Monday, April 27	Topline: S-3b by change request, E-2b as a standard change
B-1b	Monday, May 11	Biomarker pipeline revalidation
R-3	Monday, May 18	The LIMS release, with operational qualification regression
R-4	Monday, June 15	The MES maintenance window after the periodic review
R-2	Monday, June 22	After the post-topline analyses, by change request

Five changes will not happen on the roadmap's dates: four were declined at the readout and one was approved and then reversed by the CISO.

Declined and reversed changes, with the owner's own reason. R-5 and B-2 are quoted. L-4 was approved at the readout, so it counts toward the gate at its estimate; the others count for nothing.

ID	Change	Decided by	Reason	Annualized, net
C-2	EC2 Instance Savings Plan for the four remaining p4d.24xlarge, one year, No Upfront	Mateo Varga and Rafael Quintero	The training fleet moves to a newer accelerator within the year, and a family-locked plan would strand. Revisit when the new accelerator is chosen.	\$264,705
D-2	Multi-AZ off on the LIMS and MES val databases	Dr. Lorraine Osei	The val environments' performance qualification includes failover tests, and val must match production's configuration. Not revisited: the val configuration must match production.	\$80,763
R-5	Clinical data repository: 8 to 6 ra3.4xlarge nodes	Ines Halvorsen	"The CDR feeds unblinding. Nobody touches it until topline, and after topline it grows for the integrated summaries." Revisit in the fall, after the integrated summaries.	\$53,688
B-2	Genomics Batch pinned to the Lustre file system's Availability Zone (cross-AZ transfer down 80%)	Priya Okonkwo	"One zone means a third of the Spot pools. I'd rather pay the cross-AZ toll than eat interruptions." Revisit if Spot interruptions in the zone fall.	\$36,962
L-4	CloudTrail S3 data events: write events only on research buckets; read and write kept on GxP buckets	Rosalind Tran	Rosalind Tran reversed it on March 11, before it was applied: the read events on research buckets feed the insider-risk detection for sequence designs. Not revisited: the detection it feeds stays.	\$35,194

Honest limits, as of March 20, 2026

- **Nothing is measured.** Every dollar is an estimate, net of the 6% discount. The week-6 measurement uses the 7 days before and after each change, on data through March 30, 2026; a change made after March 22, 2026 would not have 7 full days after it. All four of this week's changes were in place by March 19.
- **B-1 rests on Spot prices and a retry rate.** It uses the 30-day mean and a 12% allowance. March prices and real retries will differ, and the measurement will say by how much.

- **F-1 is sized on a 90-day peak.** A workload that is not in the history, such as the long-read sequencing core planned for June, needs its own throughput review before it lands on the new file system. GR-6 builds that review in week 5.
- **S-3a depends on the hold tag.** It saves \$87,540 a year net only while the val environments run 300 hours a month; an executed protocol that holds an environment on costs part of the saving, and that is the trade QA asked for.
- **The Savings Plan estimate holds only while the floor holds.** It is measured by the plan's ARN in week 6: its fee, the usage it covered, and its utilization.
- **S3-4 is in place and counts for \$0.** It moves objects only after 30 days without access, so its first measurement window runs May 11 to May 17, 2026, with the result on May 19.
- **Enterprise Support and the Oracle licenses R-3 will free are not counted.** Support falls with any saving, and licenses are not on the AWS bill.

What comes next

Week 5 carries no savings by design. It builds what keeps the bill down and lands planned growth right: nine guardrails, GR-1 to GR-9, each with one owner, the data-class lifecycle standard, and the deferred GxP change requests going to the board on Tuesday, March 24. What the guardrails prevent in planned growth is cost avoidance, and it is **not measured, not counted toward the gate, and not a saving on the bill**. The GMP campaign freeze ends on Friday, March 27. Week 6 measures every change on Copperwick's Cost and Usage Report.

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

AWS, Amazon Web Services, AWS Graviton, and the AWS service names are trademarks of Amazon.com, Inc. or its affiliates. Oracle and Java are registered trademarks of Oracle Corporation and its affiliates. Microsoft, Windows, and SQL Server are trademarks of the Microsoft group of companies. Red Hat and Red Hat Enterprise Linux are trademarks of Red Hat, Inc. PostgreSQL is a registered trademark of the PostgreSQL Community Association of Canada. Kubernetes is a registered trademark of The Linux Foundation. Nextflow is a trademark of Seqlabs. Vantalect is not affiliated with or endorsed by any of them. Copperwick Therapeutics, Inc. is fictional.