CLOUD & AI COST GOVERNANCE DIAGNOSTIC

Sample diagnostic pack
SynthoSaaS (synthetic estate)

Fixed-fee illustrative deliverable for CTO / CFO buyers. Shows ownership, waste, safe cuts, GPU/LLM spend, and multi-cloud gaps using one fictional company. Structure matches a real engagement.

SynthoSaaS	USD 186,400	AWS + Azure
Series B SaaS - fictional	Synthetic Aug 2026 bill	EKS / GPU / LLM APIs

Pack contents (7 sections)

01 Executive summary - buyer answers first
02 Evidence register - what was reviewed
03 Cost ownership map - who owns each dollar
04 Allocation-gap table - contested / unallocated spend
05 Risk and confidence matrix - what is safe to cut
06 Act / Investigate / Hold register - decisions with owners
07 30-day action plan - client-owned next moves

Currency note: All dollar amounts are synthetic USD for one illustrative billing period (1-31 Aug 2026). Not benchmarks, forecasts, or savings guarantees.

Label: Synthetic data, real structure.

Fictional estate snapshot (SynthoSaaS)

Headcount band	~180 FTE	Product-led SaaS; eng-heavy
Primary workloads	API + data plane	EKS-hosted microservices
ML footprint	GPU pool + LLM APIs	23% of Aug bill combined
Forecast discipline	Monthly	MoM +12% triggered diagnostic
Tagging maturity	Partial	Azure stronger than shared EKS
Buyer sponsor	CTO + CFO joint	Ownership before cuts

01 EXECUTIVE SUMMARY

What finance and engineering should see first

Decision question: Who owns the month-over-month cloud and GPU rise before SynthoSaaS cuts anything?

Synthetic MoM spend rose about 12%. Growth concentrates in shared EKS, GPU retention after experiments, and LLM inference without a clear product owner. Checkout/API compute is owned and actionable. Production downsizing without SLO evidence stays on Hold.

Reconciled bill	USD 186,400	Single-period baseline for ownership work
Assigned ownership	61%	Bankable only where owner + confidence exist
Shared / contested	24%	Investigate before forecasting cuts
Unallocated gap	15%	Name the gap - do not average it away
Register posture	4 Act / 4 Inv / 3 Hold	Hold stays visible under pressure

Shared EKS growth	+USD 3,100	E-02, E-03
GPU retention	+USD 2,400	E-04
LLM inference	+USD 1,800	E-05
Non-prod idle	+USD 1,200	E-01
Product API (owned)	+USD 900	E-02
Networking egress	+USD 650	E-02

Where waste shows up

- Non-prod overnight idle without a product owner
- Orphaned GPU volumes after experiment end dates
- Sandbox accounts without budget caps

30-day posture

- Act on high-confidence idle waste
- Investigate EKS split rules and LLM token owners
- Hold prod/GPU cuts until evidence holds

Executive + registers	CTO / CFO	Tables + short narrative
Ownership + gap tables	FinOps + platform	Bankable vs contested USD
Act / Inv / Hold	Engineering leads	Owner + evidence ID per row
30-day plan	Client PMO	Client-owned milestones only

02 EVIDENCE REGISTER

What was reviewed (synthetic exports only)

No production credentials. Evidence IDs are fictional labels showing how a real pack cites sources so finance and engineering argue from the same record.

E-01	AWS CUR export (synthetic)	All linked accounts	Aug 2026	High
E-02	Cost Explorer by service	MoM delta view	Jul-Aug 2026	High
E-03	EKS namespace label dump	Shared platform cluster	2026-08-28	Med
E-04	GPU job metadata CSV	Training pool	Aug 2026	Med
E-05	LLM API vendor invoice	Inference tokens	Aug 2026	High
E-06	Azure subscription export	Secondary region apps	Aug 2026	Med
E-07	Owner roster (notes)	Product / platform / ML	Intake	Med

FinOps note: Tools can show the bill. This register shows which exports underpin each decision.

CFO	Bankable vs contested USD	Exec summary + ownership map
CTO	Blast radius of Acts	Risk matrix + Hold register
Platform lead	EKS + sandbox gaps	Allocation-gap table
ML Ops	GPU idle policy	Ownership map + register A3/H2
FinOps analyst	Evidence trail	Evidence register + Act rows

Single billing period lock	All USD rows	Avoid MoM noise in ownership %
Evidence ID on every Act row	Register section	Audit trail for finance sign-off
No credential collection in sample	This PDF	Real pack uses client-owned exports
Hold list is published	Week 1 governance	Prevents silent scope creep on cuts

03 COST OWNERSHIP MAP

Who owns every material dollar

Synthetic multi-cloud estate: AWS-primary (~88%) + Azure secondary (~12%). Totals reconcile to USD 186,400.

Product API - prod compute	AWS	Assigned / Product Eng	52,100	28%	Assigned
Customer data plane	AWS	Assigned / Data	31,400	17%	Assigned
Shared EKS platform	AWS	Shared / Platform	28,900	16%	Contested
GPU training pool	AWS	Shared / ML Ops	24,700	13%	Policy gap
LLM inference APIs	Vendor	Assigned / AI Prod	18,200	10%	Owner thin
Non-prod + sandbox	AWS	Unallocated gap	16,800	9%	Gap
Networking / egress	AWS	Shared / Platform	8,600	5%	Investigate
Azure app / misc	Azure	Assigned / Product	5,700	3%	Assigned

Buyer answers this map unlocks

Ownership: 61% named; 39% shared or gap - not one ’cloud waste’ number.
GPU / AI spend: Training + LLM APIs = 23% of bill; idle policy and token ownership are blockers.
Multi-cloud: Azure is small but tagged; AWS shared platform is where attribution breaks.

Assigned	Forecastable with named owner	Track in unit economics
Contested	USD real; split rule missing	Investigate before target setting
Gap	No owner class	Name owner in intake workshop
Policy gap	Technical owner; policy thin	ML/platform policy ticket

04 ALLOCATION-GAP TABLE

Contested and unallocated spend - named, not averaged

EKS namespaces w/o chargeback	14,200	3 products share ingress; no split	Platform + FinOps	E-03
GPU nodes past experiment end	9,800	Jobs owned; idle policy is not	ML Ops	E-04
Non-prod overnight idle	7,400	No product accountable for schedule	Platform Eng	E-01
LLM tokens without SKU map	6,100	Invoice exists; commercial owner thin	AI Prod + Fin	E-05
Sandbox accounts uncapped	4,900	Spend drifts outside forecast	FinOps	E-01
Egress spike - service unknown	3,200	Cannot assign without path attribution	Networking	E-02

FinOps vs diagnostic: A FinOps tool surfaces anomalies. This table decides which dollars are not yet bankable until an owner and rule exist.

05 RISK & CONFIDENCE MATRIX

What is safe to cut - and what is not

Act only when confidence is high and residual risk is low. Hold is a first-class outcome.

Stop idle non-prod nights	High	Low	Yes - Act	Clear schedule gap; easy rollback
Rightsize oversized API ASGs	High	Low	Yes - Act	Utilization evidence; Product-owned
Delete orphaned GPU volumes	High	Low	Yes - Act	Past end date; no active jobs
Cap sandbox budgets	Med	Low	Yes - Act	Policy change; low blast radius
Allocate shared EKS costs	Med	Med	No - Investigate	Needs split rule before banking
Trace egress to service	Med	Med	No - Investigate	Attribution incomplete
Map LLM tokens to SKU	Med	Med	No - Investigate	Owner roster incomplete
Buy RIs / Savings Plans now	Low	High	No - Hold	Commitment before ownership clear
Prod downsize without SLO	Low	High	No - Hold	Reliability risk > bill optics
Collapse multi-AZ for savings	Low	High	No - Hold	Architecture / DR risk

High	Export + owner agree	Act candidates A1-A3
Med	Export solid; owner/rule partial	Investigate I1-I3
Low	Commitment or SLO unknown	Hold H1-H3 until clarity

06 ACT / INVESTIGATE / HOLD REGISTER

Decisions with confidence, risk, and owner

A1	Stop idle non-prod nights	High	Low	Platform Eng	Act	E-01
A2	Rightsize oversized API ASGs	High	Low	Product Eng	Act	E-02
A3	Delete orphaned GPU volumes	High	Low	ML Ops	Act	E-04
A4	Cap sandbox account budgets	Med	Low	FinOps	Act	E-01
I1	Allocate shared EKS costs	Med	Med	Platform + Fin	Investigate	E-03
I2	Trace egress spike to service	Med	Med	Networking	Investigate	E-02
I3	LLM token owner unclear	Med	Med	AI Prod + Fin	Investigate	E-05
I4	Spot interruption impact	Low	Med	Platform Eng	Investigate	E-01
H1	RI / SP purchase now	Low	High	FinOps	Hold	E-01
H2	Prod downsize without SLO	Low	High	SRE	Hold	E-07
H3	Multi-AZ collapse proposal	Low	High	Architecture	Hold	E-07

Illustrative near-term Act candidates (not a savings promise): ASG floors ~USD 28-36k/yr teaching range; orphan cleanup ~USD 9-14k; non-prod idle ~USD 12-18k. Real estates differ.

Act	4	~USD 42-68k annualized equiv.	Include after owner sign-off
Investigate	4	~USD 38-51k contested	Exclude from forecast
Hold	3	Commitment / SLO exposure	Disclose; no target

07 30-DAY ACTION PLAN

What the client owns next

Implementation stays with SynthoSaaS. AICloudStrategist does not execute production changes as part of the diagnostic fee.

Days 1-7 Confirm owners

Lock owners for Act items A1-A4. Freeze the Hold list so commercial pressure cannot erase H1-H3. Name the EKS split-rule owner.

Days 8-14 Execute high-confidence Acts

Run A1/A3 in non-prod with rollback notes. Open Investigate tickets I1-I3. No production GPU changes.

Days 15-21 Investigate shared platform

Draft EKS namespace/ingress allocation options. Keep contested dollars out of finance forecasts until the rule is agreed.

Days 22-30 Hold review + re-measure

GPU idle policy workshop. Re-measure baseline. Promote Investigate to Act only where evidence and owners now exist.

1	Signed owner list A1-A4	No anonymous infra spend in forecast
2	Rollback notes for A1/A3	Change tickets linked to evidence
3	EKS split options memo	Finance stops allocating contested 16%
4	Re-measured Aug-equivalent	Investigate rows promoted or closed

What we typically need from the client

Billing export / CUR-equivalent	Reconcile ownership to one period	After fit, under NDA
Account list + known product owners	Seed the ownership map	Intake
Optional: labels, GPU jobs, model bills	GPU / LLM attribution	If in scope

Act owners confirmed	Client PMO	A1-A4 named in writing
Investigate tickets opened	Platform + Fin	I1-I3 in backlog with dates
Hold list ratified	CTO / CFO	H1-H3 unchanged in forecast deck
Baseline re-measure	FinOps	Same period definition as diagnostic

BOUNDARIES & NEXT STEP

Synthetic data, real structure

What this diagnostic does NOT do

- No production login or credentials required to start
- No savings-share contract and no savings guarantee
- No implementation of cuts inside the diagnostic fee
- Not a statutory financial audit or certification

Primary CTA

Request a diagnostic fit check

Free suitability call. About two minutes on the form. No billing files required. We respond within one business day (IST).
https://aicloudstrategist.com/request-fit-review/

SynthoSaaS is fictional. This pack demonstrates deliverable shape for fit-review buyers - not a client endorsement, case study, or promised outcome.

AICLOUDSTRATEGIST PRIVATE LIMITED | contact@aicloudstrategist.com | +91 87963 02608