CB CloudByte PMS

ENTERPRISE BRIEFING · CIO / CFO / VP ENGINEERING

# Every AI-assisted change in your codebase traced end-to-end, unified into one knowledge base, costed to the token.

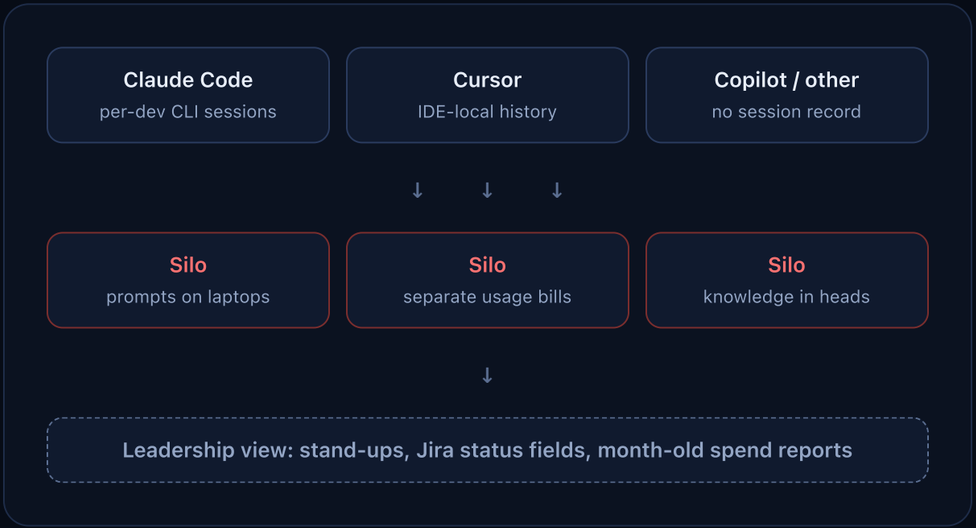
CloudByte PMS is the execution-visibility and governance layer for AI coding agents - Claude Code, Cursor, and beyond - built for the leaders accountable for delivery, spend, and risk.

MULTI-PLATFORM CAPTURE | END-TO-END TRACING | COST TRACKING & OPTIMIZATION | UNIFIED KNOWLEDGE BASE

WHY NOW

# AI agents now write a majority of your code. Your visibility stack was built for humans.

Each agent ships its own console, its own logs, its own billing. Nothing rolls up. The people accountable for the output are the last to see it.




- **70-90% of new code** in AI-forward teams is agent-assisted - with no system of record for how it happened.
- **Per-seat AI spend is scaling faster than headcount** - and lands on the CFO's desk as an unexplained line item.
- **Auditors and security teams are starting to ask** "which AI wrote this change, and from what instruction?"
- The window to put governance in *before* the tooling sprawls is now - not after standardization is impossible.

Range reflects commonly cited estimates for AI-forward teams — ask us for sourcing, or we'll benchmark your org directly during the pilot.

THE PROBLEM

# Prompts, observations, and knowledge are segregated - so you can't manage what you can't see.

Three silos, three owners, zero correlation. Every question below currently takes a meeting to answer.

### Prompts - invisible

- Instructions to AI agents never leave the developer's machine
- No record of *what was asked* behind any change
- "Why was this file touched?" = tribal memory

### Observations - unclassified

- Work happens as bugfixes, features, refactors, decisions - recorded nowhere
- Status reporting is manual, weekly, and already stale
- Managers rediscover blockers at the next stand-up

### Knowledge - evaporating

- Every fix produces reusable insight; none of it is captured
- Same problems re-solved by different devs, weeks apart
- Attrition = knowledge walks out the door

**5–8 hrs**

per manager, per week, reconstructing status across tools

**0%**

of AI-agent prompts & reasoning retained today for audit or reuse

**?**

AI spend per project, per developer, per outcome - currently unanswerable

"0%" reflects the typical starting state we see pre-engagement — no org we've onboarded had prompt/reasoning retention in place before CloudByte PMS.
5–8 hrs is our estimate for a manager overseeing 5–8 engineers across Jira, Git, and Slack — based on typical status-reconstruction patterns we've observed; your pilot will measure your own number.

COST OF INACTION

# The status quo doesn't hold - it compounds.

Modeled for a 50-engineer org with typical AI-agent adoption.
Your numbers, calibrated live during the pilot.

**~$380K/yr**

manager + PMO hours lost to manual status reconstruction (50 eng org)

**20–30%**

duplicated AI spend - same problems re-prompted across the team

**4–6 wks**

onboarding time that captured knowledge could cut in half

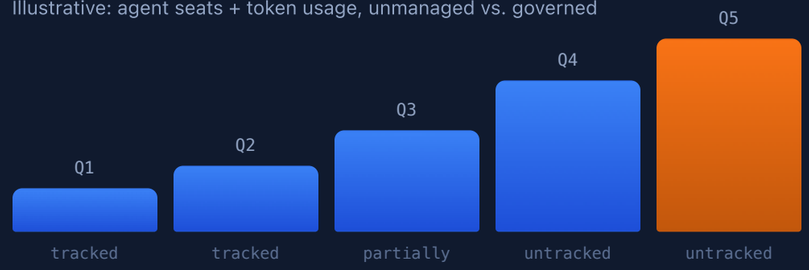
**1 audit**

Exposes unremediated AI-change gaps - provenance can't be reconstructed after the fact.

ROI model & assumptions in appendix — every input adjustable to your org.

### Untracked AI spend, compounding

Illustrative: agent seats + token usage, unmanaged vs. governed

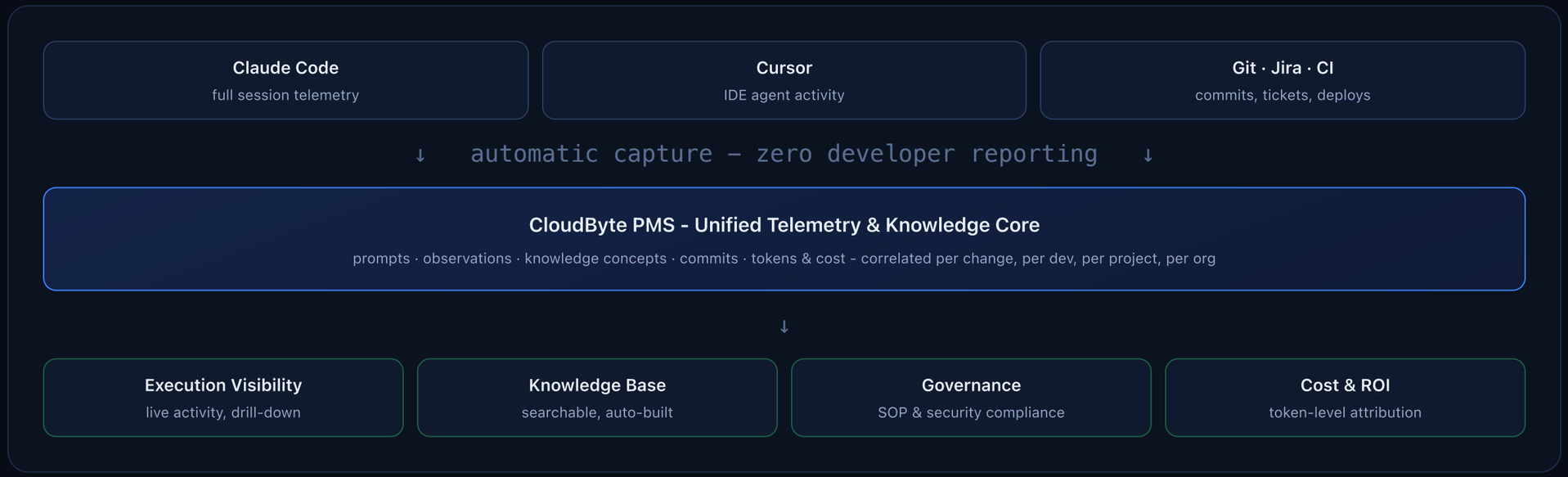



Spend grows with adoption either way. The only question is whether you can see it, attribute it, and optimize it.

THE SOLUTION

# One platform. Every agent, every prompt, every change - unified, traced, and governed.

CloudByte PMS sits beside your developers' AI tools and turns their raw activity into a single, org-scoped system of record.



HOW IT WORKS

# Four steps. Zero reporting effort from developers.

A lightweight sync agent installs in minutes per developer. Everything after that is automatic.

1

### Capture

Sync agents record every AI session - prompts, tool calls, files read and changed, tokens consumed - from Claude Code, Cursor, and Git.

→ 2

### Classify

Each unit of work is auto-classified: bugfix, feature, refactor, change, discovery, decision - with a narrative, observed facts, and knowledge concepts.

→ 3

### Unify

Everything correlates in one org-scoped core: prompt → reasoning → commit → ticket → deploy → cost. One trace per change.

→ 4

### Govern

Dashboards, DORA metrics, SOP compliance scoring, secret scanning, and cost analytics - scoped by role, isolated by org.

**< 15 min**

to install per developer - no code changes, no workflow changes

**100%**

of AI-assisted work captured automatically at the source

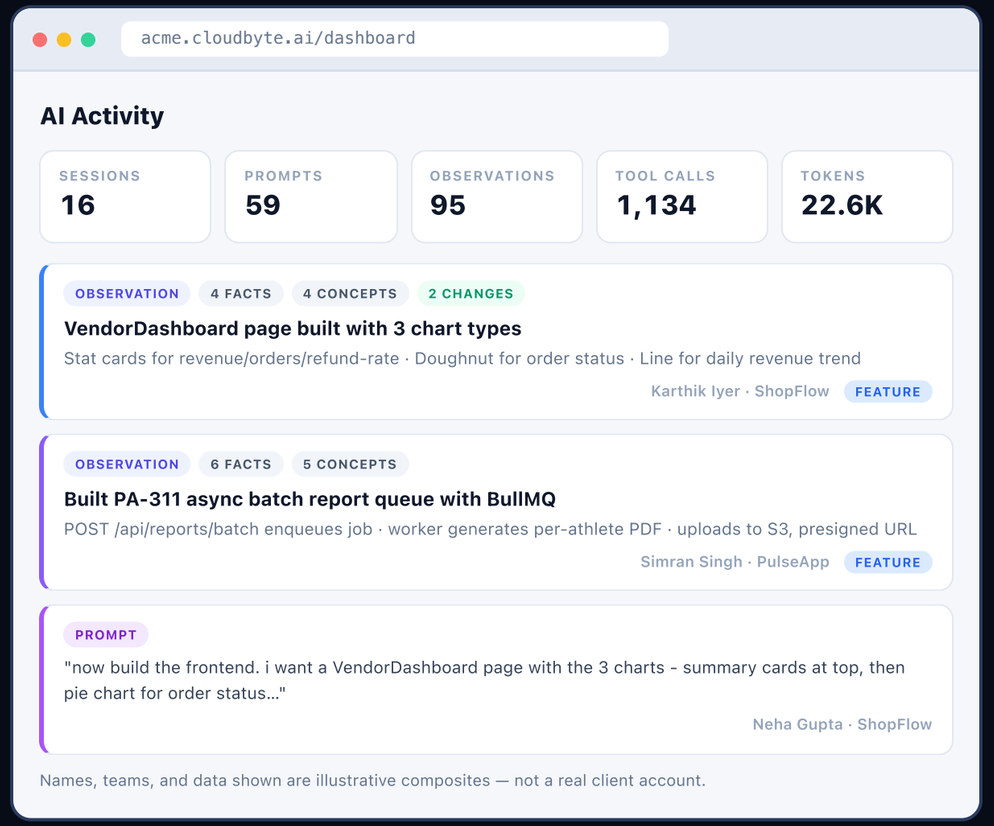
**1 trace**

from prompt to production for every change

THE PRODUCT – LIVE VIEW

# What used to surface in Friday's status meeting is on screen the moment it happens.

- **AI Activity dashboard** - sessions, prompts, observations, tool calls, tokens, rolled up per org / project / user / date.
- **Observation feed** - every piece of work classified, with facts, knowledge concepts, and source changes attached.
- **Full drill-down** - from a headline to the exact prompt, reasoning, and files touched.
- **No self-reporting.** Nobody filled in a form to produce this screen.



KEY CAPABILITIES → YOUR OUTCOMES

# Built for the three people in this room.

### VP Engineering - delivery truth

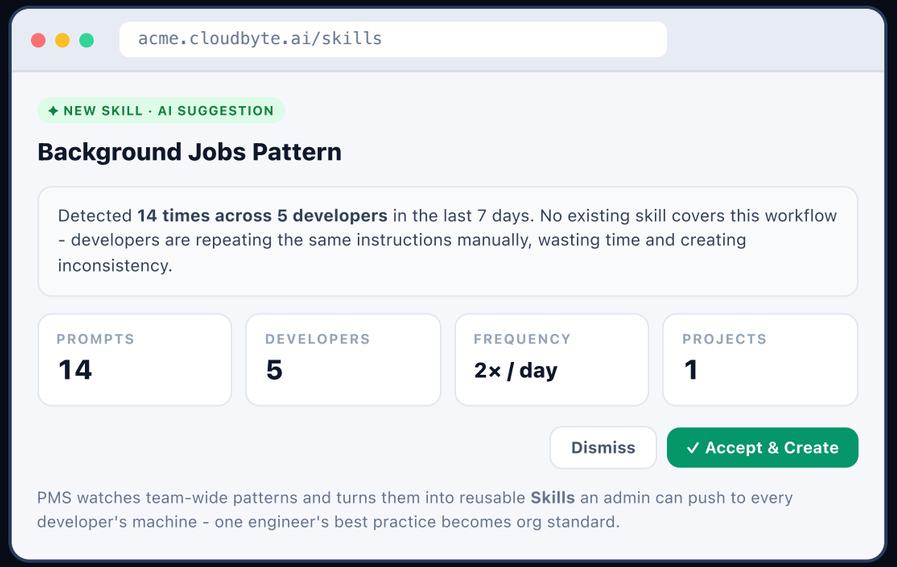
DORA metrics, value-stream p50/p75/p90, stuck-session detection, developer insights for 1:1s - computed from real activity, not self-reports.

### CFO - cost you can attribute

Token-level cost tracking per model, per developer, per project. Hours-saved ROI, adoption rates, and duplicate-spend detection - the AI line item, explained.

### CIO - governance & audit-readiness

SOP compliance scoring, secret scanning before leak, RBAC + per-org tenant isolation, and a full provenance trail for every AI-assisted change.



DIFFERENTIATION

# The honest comparison - including "do nothing."

| | CLOUDBYTE PMS | PER-TOOL CONSOLES (CLAUDE / CURSOR ADMIN) | JIRA / PMO SUITES | DO NOTHING |
|---|---|---|---|---|
| Multi-platform AI capture | ✓ **Claude Code + Cursor, one view** | ✕ one vendor each | ✕ none | ✕ |
| End-to-end tracing<br>prompt → reasoning → commit → deploy | ✓ **one trace per change** | △ usage stats only | △ ticket status only | ✕ |
| Cost tracking & optimization | ✓ **token-level, per dev / project / model** | △ aggregate bill | ✕ | ✕ invisible spend |
| Unified knowledge base | ✓ **auto-extracted concepts, searchable** | ✕ | ✕ manual wikis | ✕ tribal memory |
| Governance & SOP compliance | ✓ **continuous scoring** | ✕ | △ manual audits | ✕ audit exposure |
| Developer effort required | **Zero reporting** | **Zero** | High - manual updates | **Zero** |

Where a competitor is genuinely fine, we say so. The gap is the correlation layer — nobody else joins prompts, changes, knowledge, and cost in one record.

PROOF & ROI LOGIC

INTERNAL PILOT DATA

# From our pilot org: visibility you can act on, savings you can defend.

**6–9 hrs**

manager time saved / week - status & 1:1 prep auto-generated

**89.3 hrs**

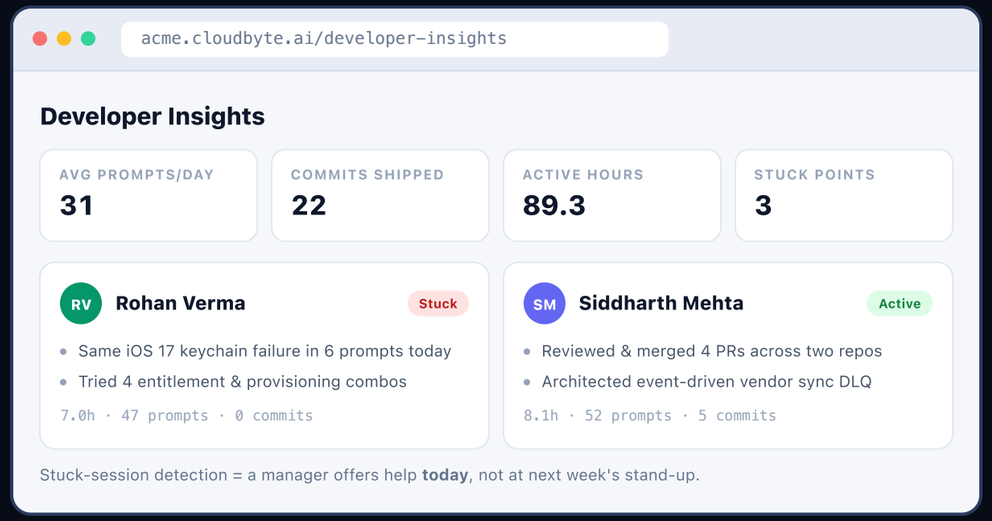
active engineering hours attributed in one day, pilot org

**15–25%**

token-spend reduction target from de-duplication & skills reuse

**3 → 0**

stuck points resolved same-day once surfaced automatically

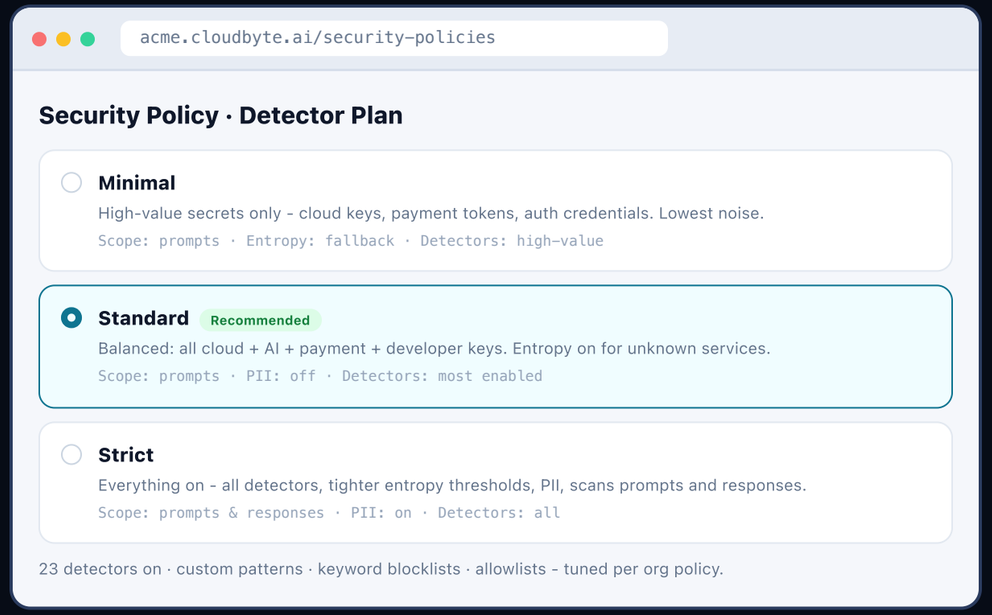



- **ROI logic:** (manager hrs saved + duplicate work avoided + AI spend optimized) vs. per-seat price - payback modeled inside the product, on your data.
- Pilot exits with a written ROI report your CFO can audit.

Figures from CloudByte internal pilot org (shown) — your pilot produces your own.

SECURITY & TRUST

# Governance strong enough to make AI coding safe to scale.




- **Secrets stopped before they leave** - keys, tokens, credentials detected in prompts *before* they reach any model or commit.
- **Hard multi-tenant isolation** - every query org-scoped at the API layer; subdomain-per-org; no cross-tenant leakage by design.
- **RBAC with four roles** - super admin → org admin → reviewer → standard; server-enforced, never client-trusted.
- **Google SSO**, encrypted in transit and at rest, hosted on AWS (Mumbai region; region options on enterprise plans).
- **Audit-ready by default** - the provenance trail doubles as your AI-change audit log.

IMPLEMENTATION

# Live in days. Proving value inside two weeks.

No code changes, no workflow changes, no data migration. IT effort: SSO approval and about an hour.

DAY 1–2

### Connect

Google SSO enabled, org tenant provisioned, sync agent installed per developer (<15 min each). Jira & Git connected.

WEEK 1

### Capture

Activity flowing automatically. First dashboards, observations, and cost data live. Security policy tuned to your standard.

WEEK 2–4

### Calibrate

DORA baselines established, SOP compliance rules mapped to your process, ROI model loaded with your actual rates.

MONTH 2–3

### Prove

Full pilot review: hours saved, spend optimized, governance coverage - written ROI report and rollout plan.

**~1 hr**

total IT effort for setup

**0**

changes to developer workflow

**Days**

to first live dashboard, not weeks

THE ASK

# One team. Three months. Free. Measured against numbers we agree on today.

Scope: one engineering team (5–15 devs) · Success metrics: manager hours saved, AI spend attributed, governance coverage · Exit: written ROI report + rollout decision.

**Scope the pilot this week →** | **getpms.cloudbyte.ai**

## This week

30-min scoping call, team + metrics chosen

## Next week

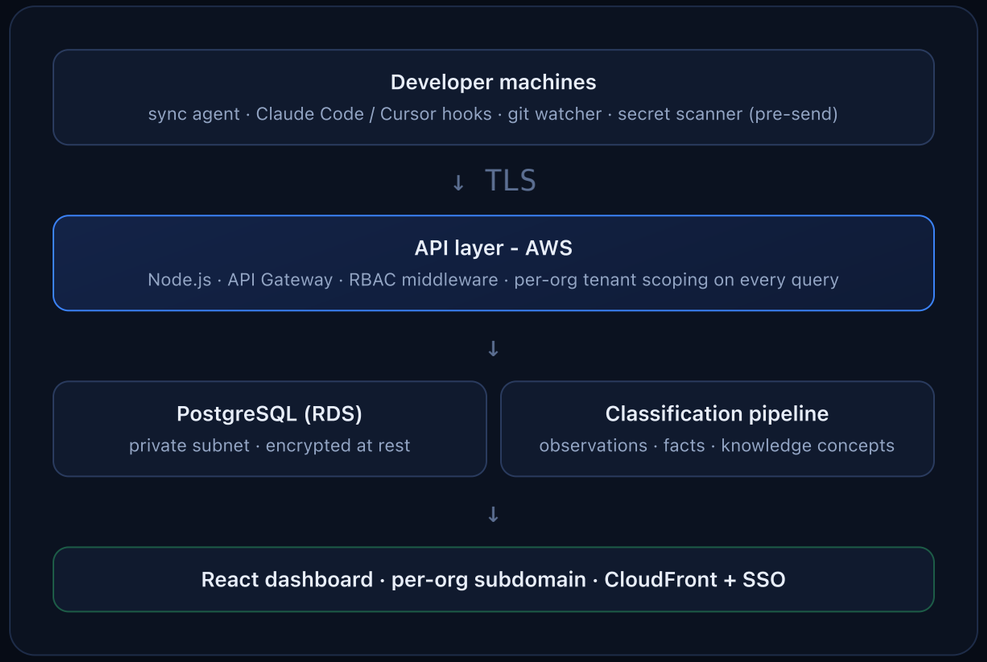
agents installed, data flowing

## Day 90

ROI review with your own numbers

APPENDIX – ARCHITECTURE & FAQ

# Technical reference




### Is this developer surveillance?

No keystrokes, no screen capture. We record work product and process - the same things a code review or stand-up covers, automated. Stuck-detection is framed and used as a support signal.

### What about our source code?

PMS stores prompts, session metadata, diffs summaries and file paths - not your full repository. Data is org-isolated and encrypted; retention windows configurable.

### Which tools are supported today?

Claude Code (deep session telemetry) and Cursor, plus Git (local & remote), Jira, and CI signals for DORA. Additional agent platforms land on the same unified schema.

### Can we self-host / choose region?

SaaS runs on AWS ap-south-1 today; dedicated-region and private-deployment options are enterprise-tier conversations.