COPEBIT

Containers & AWS Meetup · Zürich · 03.11.2026

Enterprise Claude Code & Cowork, built natively on AWS

What the SaaS enterprise plan gives you,
and what it takes to build it yourself.



Roman Frei · Product Owner, copebit AG

copebit is an AWS Generative AI Competency Partner and an AWS Certified Anthropic Resell Partner

Engineers **“We need Claude!”**

CISO **“Where does our data go?”**

Not a proof of concept

This runs in production today

100 → **300**

developers today — next

Engineers wanted Claude.

Security and compliance needed more than the standard plan offers.

The project in one table

The goal: every SaaS feature, plus compliance

	Claude Enterprise Anthropic SaaS	**Built on Amazon Bedrock** in your own cloud account
What the teams need		
Company login (SSO)	✓ included	✓ we run it
Spend limits per user	✓ included	✓ we run it
Central settings	✓ included	✓ we run it
Tool connectors (MCP)	✓ included	✓ we run it
Shared skills & knowledge	✓ included	✓ we run it
Usage reporting	✓ included	✓ we run it
What compliance asked for		
Inference only in Europe	✗ missing	✓ cloud feature
Private network, no internet	✗ missing	✓ cloud feature
Data Loss Prevention	! bring your own	✓ cloud feature
Every prompt in our storage	✗ missing	✓ cloud feature

How it fits together

Two separate paths, one audit trail

Model path

your prompts & answers
private · stays in Europe

Engineer
Claude Code · Cowork

→

Zero Trust access
company network

→

Model Gateway
login · limits · settings

→

Gateway Proxy
checks every request

→

Claude on Bedrock
Europe regions

Tool path

only what a tool needs,
e.g. a search term

Engineer
company login

→

Enterprise MCP Gateway
one door for all tools

→

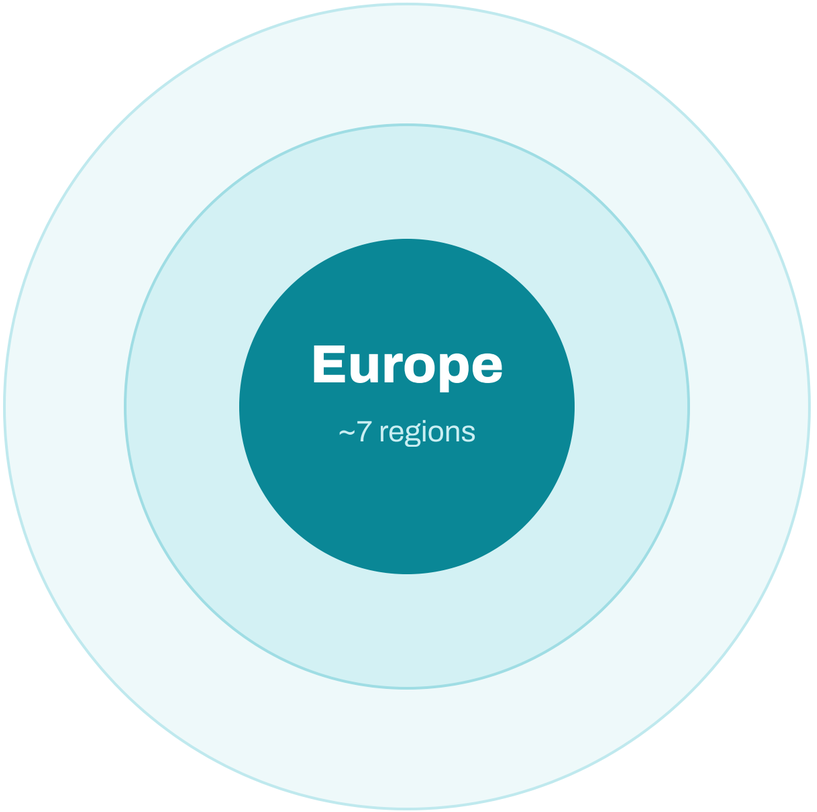
- Web search
- SaaS connectors
- Internal tools
- Skills Registry
- Knowledge DB

Reporting & audit

both paths write here

every prompt & answer · every tool call · every cent spent

Europe only, enforced in three places



- **1 · Gateway**
  only offers models that run in Europe
- **2 · Cloud permissions**
  the gateway may only call models in Europe
- **3 · Cloud-wide policy**
  blocks every non-Europe model, for every account

There is no “Zurich only”. Requests spread over ~7 European regions, so every control has to exist in each of them.

Model calls never touch the public internet



Same pattern as your corporate VPN

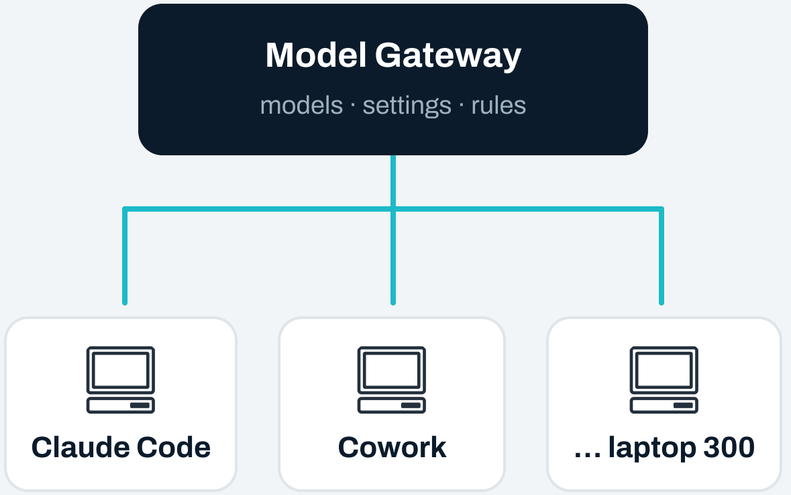
Existing access policies and audit reports apply. Nothing new for security to learn.

No long-lived keys, anywhere

Credentials are short-lived. Even a valid session can't reach the model from outside the network.

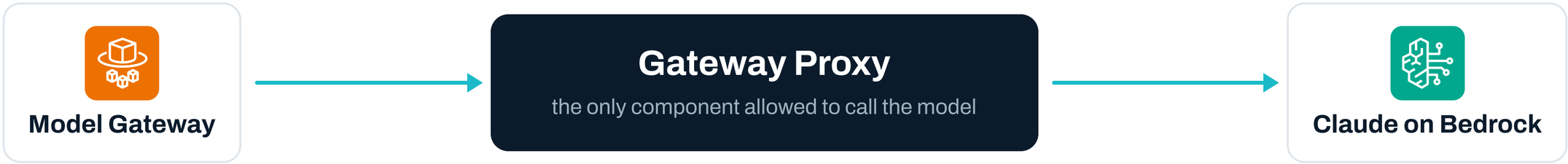
One front door to the model

- **Log in with Microsoft Entra ID**
  `/login` once, no cloud account, no keys on the laptop
- **Roles come from Entra groups**
  a new developer is in the right group, so they're ready on day one
- **Personal spend limit, checked live**
  see it with `/usage`
- **Central settings for Claude Code & Cowork**
  change once, not on 300 laptops



Device management only sets **the gateway address**.
Everything else comes from the gateway.

A checkpoint on every request



Guardrails

block personal data, passwords, harmful content

Pattern filters

your own terms: customer names, key formats

Mask, don't block

hide the sensitive part, the session continues

Rules per team

stricter for teams with sensitive data

Fallback

switch region when one is overloaded

Enterprise MCP Gateway: one door for every tool

Without it, tools run on every laptop

- Passwords for SaaS tools in local files
- 20 versions of the same tool
- Nobody sees who used what

With one gateway

- **Credential vault**
  tool passwords stored centrally, never on laptops
- **Rules on every call**
  default deny: who may use which tool, every decision logged
- **Managed web search**
  admins pick the allowed sites, the model can't change that
- **Internal tools hosted once**
  one version for everyone

Semantic tool search keeps the context window free

Without every connected tool's description rides along with every prompt



With semantic search the agent asks the gateway “which tools fit this task?”



Fewer tokens per prompt, and adding the 100th tool doesn't make every prompt more expensive.

Proportions illustrative

Share what works: skills and knowledge

Skills Registry

- **One catalog**
  skills, tools (MCP servers) and finished agents
- **Works with any agent**
  Claude, Codex, Kiro and others
- **Agents look things up themselves**
  nothing pre-installed on every laptop

Knowledge DB

- **One knowledge base the agent can search**
  fed from wiki, SharePoint, file storage
- **You only find what you may read**
  existing permissions apply
- **Stays in your storage**
  deleting data is your decision

What engineers see

For the user, it's just Claude

1

Sign in once

with the company Microsoft account

2

Get what the group allows

- models
- tools
- skills & knowledge
- a personal spend limit

3

Work as usual

in the normal Claude Code and Cowork

> Everything from the last slides runs **behind the scenes**. The engineer never sees it.

Who asked what, when, with which model



Every prompt & answer

kept 365 days, in your own storage



Find any prompt

by person, date or model, with one query

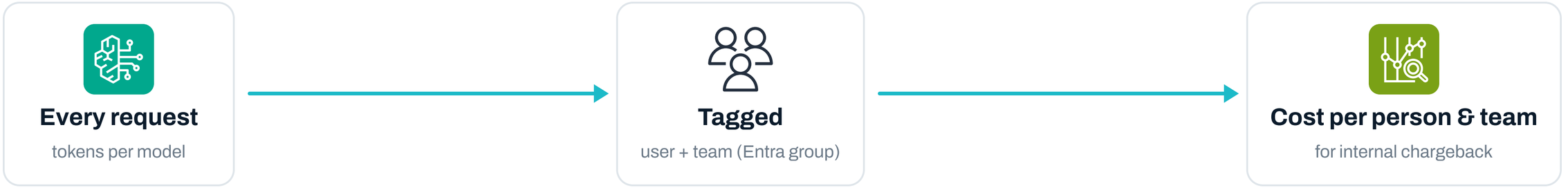


Every tool call

who used which tool, allowed or denied

> Audit questions get answered **without asking a vendor.**

Every cent has a name on it



Weekly cost email

who used how much, per engineer

Budget alerts

warn early, stop at the limit

Your own control center

costs, limits and tool rules in one screen

The honest part

It isn't cheap

1

Build

Months of platform engineering before the first engineer logs in.

2

Maintain

New models, new Claude Code releases, AWS services still in preview. Something changes every week.

3

Run

Always-on infrastructure and logs, plus every token at pay-per-use prices.

> If your compliance doesn't need it, **buy the SaaS plan.**

So, which one?

SaaS or self-built?

Compliance demands Europe-only, private, every prompt logged

Build it on Bedrock

Plan for the engineering, and for running it.

Otherwise

Buy the SaaS plan

More features, ready today, and the vendor runs it.

Four things to take home

1 Ask compliance first

Collect their questions, then build exactly what answers them.

2 Never rely on one setting

Block important things in two or three places.

3 The model is the easy part

Connecting tools safely takes most of the work.

4 Send everything through your gateway

That's where you control, filter and measure.

Questions?

Roman Frei

Product Owner, copebit AG

roman.frei@copebit.ch

linkedin.com/in/roman-frei

connect on LinkedIn