

Hybrid Infrastructure

Questions this doc answers

- Can some teams use cloud AI tools while others stay on private models?
- How do private and hosted shapes coexist conceptually?
- What should we validate in a hybrid pilot?

Shared memory across hybrid stacks

Reflect separates the **memory layer** from the **model layer**. Teams can mix:

- Hosted or isolated Reflect cloud for general work, and
- Private/self-host Reflect for regulated data planes,

...as long as identity, sharing policy, and residency requirements are designed up front. Both sides are still "Reflect Memory" — the difference is where the store and network boundary live.

Common patterns

1. **Cloud-first evaluation, private production** — Start hosted to prove workflow, then move regulated workloads to isolated or self-host.
2. **On-prem models + private Reflect** — Keep models and memory inside the VPC; connect only approved corporate AI clients.
3. **Split by data class** — Non-sensitive team knowledge in hosted; PHI/PCI-class context in private.

Guardrails to confirm in pilot

- Which data classes may leave the private boundary
- Whether public webhooks/integrations are allowed
- SSO and audit export requirements
- Which AI clients are approved to connect

Networking recipes and allowlist configuration ship with the private deploy package under NDA.