

What is the VividCloud Agentic AI Accelerator?

The Agentic AI Accelerator is VividCloud's proprietary framework for building you a modular, AWS-native agentic system — one central **Planner agent** and a team of specialized AI agents that work across your organization's operational data, external APIs, documents, and telemetry.

Unlike a single chatbot, the Planner decides what's actually needed to answer each request and orchestrates the right specialists in real time — looping through a 5-step context cap until the answer is complete, with every step streamed back visibly rather than hidden behind the scenes.

Who owns the IP, and what does this cost?

Because the Accelerator builds a custom system rather than licensing a product, **you own all the resulting IP** — there are no product costs and no ongoing royalties. You pay for the engagement that designs, builds, and deploys your system; what comes out of it is yours.

How does it differ from a traditional RAG (Retrieval-Augmented Generation) solution?

Traditional RAG systems answer questions based on retrieved content alone.

The Accelerator goes beyond retrieval by building in agent orchestration — a Planner agent coordinates specialized agents for operational data, external APIs, document knowledge, and visualization, each of which reasons over data, validates results, and can trigger downstream actions.

This enables context-aware decisions and dynamic automation, rather than a static information lookup.

How responsive is the system?

The Planner streams its reasoning back as it works, rather than making you wait for a final answer, and loops through its 5-step context cap only as needed — so simple questions resolve quickly and complex ones stay visible while they're being worked through.

What kinds of tasks can these agents perform?

Each agent is purpose-built for a domain of expertise:

- **Operational Data Agent:** Writes, executes, and self-repairs queries against events, telemetry, and logs, then explains the results in plain language.
- **External APIs Agent:** Auto-discovers a REST API's OpenAPI spec, plans a call, executes it, and repairs failed requests with corrected parameters.
- **Document Knowledge Agent:** Hybrid vector and keyword search across your document library, reranked, cited, and automatically re-searched when the first pass falls short.
- **Visualization Agent:** Queries the data, drafts a chart, renders it, looks at the result, and redraws it until the chart actually communicates the answer.
- **Continuous Memory Agent:** Extracts durable facts from every conversation and queues them for expert review before they inform the next answer.
- **Planner (Orchestration) Agent:** Reformulates the question, selects the next agent, and decides when the answer is complete.

What's the difference between using it in chat versus calling it from our own systems?

The system offers two ways in, built on the exact same agents.

Conversational — for people who need to dig in: analysts and supervisors get an interactive chat experience that streams reasoning as it happens, not just a final answer.

Headless — for systems that need an answer: your own services integrate directly via `POST /query/stream`, an authenticated endpoint that returns the same event-streamed agent reasoning with no UI involved.

How does the system ensure accuracy, explainability, and trust?

Every response carries provenance and reasoning context, so it can be traced back to its source rather than treated as a black box.

The Continuous Memory agent queues newly extracted facts for expert review before they become part of what the system relies on, and ongoing prompt and agent-correctness testing refine performance alongside that human-in-the-loop check.

What AWS services form the core of the system?

The system is fully cloud-native and AWS-based:

- Amazon Bedrock for LLM reasoning
- Amazon Cognito for authentication and access control
- Aurora with pgvector as the knowledge base for document and fact embeddings
- Elasticsearch / OpenSearch for operational data
- Amazon DynamoDB for conversation history and state management
- ECS/Fargate and Lambda for compute, provisioned via CDK for consistent, repeatable deployment

How does the system handle security and data privacy?

Security is foundational. All components reside within your AWS environment, protected by IAM-based access control, VPC isolation, and encryption at rest and in transit.

Cognito with SSO integration ensures authenticated access, while Bedrock's managed model endpoints guarantee that private data never leaves your AWS boundary.

How does the system evolve over time?

The Continuous Memory agent extracts durable facts from every conversation and queues them for expert review rather than applying them automatically.

Once approved, those facts become part of what every agent can draw on — so the system's shared knowledge grows more aligned with your organization over time, under human control.

Can it be customized for different industries or enterprise systems?

Yes — domain configurations swap in without modifying agent code, so the same engine can be tuned to a new vertical instead of being rebuilt. Two examples are live today:

- **Robotics & Warehouse Automation:** a floor supervisor asks about downtime and gets a cited answer; a fleet-monitoring service asks the same question headlessly every 15 minutes, with a qualifying answer triggering a reroute or maintenance ticket automatically.
- **Property & Casualty Insurance:** adjusters cross-reference claims against loss history and cited guidelines, while a claims-intake system calls the same agents on First Notice of Loss to route standard processing versus fraud review.

What's the implementation and onboarding process?

VividCloud provides end-to-end deployment and tuning:

1. Environment setup in your AWS account
2. Data and API integration
3. Agent configuration and validation
4. Governance, security, and cost-control alignment
5. Training and handoff for your operational team

Our AWS-certified engineers and solution architects ensure a smooth path from pilot to production.

How do we get started with a pilot or demo?

Contact VividCloud to schedule a consultation or proof of concept. As an AWS Advanced Tier Partner, we may be able to offer AWS-funded PoCs to qualified clients.