

THAIKI Cloud

PRAXIS

Contents

1.	Customer Pain Points	04
2.	Competitive Differentiation	06
3.	Platform Positioning & Target User	08
4.	Praxis Product Composition	010
	Praxis	
	Agent Builder	
	Enterprise RAG Engine	
	MCP Tool ub	
	Chat Interface	
5.	Observability · Security · Operations	016
6.	Business Impact	018
	Appendix #1 High-Level Architecture	020

1.

Customer Pain Points

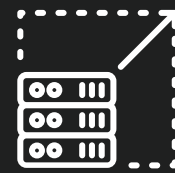
AI infrastructure operation's cost and complexity are blocking the expansion speed of enterprises.

Single vendor lock-in, workload optimization burden, operation difficulty, and data sovereignty regulations are acting as major obstacle factors for enterprise AI transition.



Data security and regulatory compliance issues

- Inability to meet closed network/on-premise requirements of financial, medical, and public institutions.



RAG quality and domain suitability issues

- Difficulty managing by integrating various data sources.
- Low accuracy of general RAG (lack of domain-specific terms, abbreviations, and context understanding).



AI Agent construction and operation complexity

- Difficulty in tracking Agent execution status, performance, and cost.

2. Competitive Differentiation

Praxis provides clear technological and operational differentiation.

Category	Competitor Strengths	Competitor Limitations	THAKI CLOUD
Agent Studio	<ul style="list-style-type: none"> • Large-scale enterprise search • Pre-built connectors 	<ul style="list-style-type: none"> • High Cost • Cloud dependency • Complex settings required 	<ul style="list-style-type: none"> • On premise support • Cost efficient
Domestic Small-scale Solutions	<ul style="list-style-type: none"> • Specialized for closed network environments 	<ul style="list-style-type: none"> • Domestic small-medium size solutions • No MCP support 	<ul style="list-style-type: none"> • MCP Native support • HTTP/STDIO types • K8s dynamic deployment • OAuth authentication integration
No-code LLM Builder	<ul style="list-style-type: none"> • No-code LLM Builder 	<ul style="list-style-type: none"> • Workflow-centric rather than Agent-centric • No MCP support 	<ul style="list-style-type: none"> • Hybrid RAG engine • Various chunking strategies

3. Platform Positioning & Target User

A secure AI agent development and deployment platform.

Provides a production-ready Praxis performing decision-making based on corporate internal data.

Target Industries

Finance

Customer consultation automation, compliance review, risk analysis

Manufacturing /Semiconductor

Technical document search, quality control analysis, facility monitoring

Public/Medical

Civil complaint processing automation, policy review support, clinical data analysis

Target Persona

Enterprise Platform Owner

Requirements

- Group/team unit management
- Operational stability and scalability
- Escaping LLM vendor lock-in

Praxis Provides

- Project-based multi-tenancy
- Claude/GPT/On-premise LLM options
- Helm-based enterprise deployment
- vLLM self-embedding operation.

SI/AI Solution Architect

Requirements

- Fast customer Agent construction
- External business system (ERP/CRM) linkage
- Reusable template

Praxis Provides

- MCP template-based tool integration
- LangGraph Think→Act→Observe cycle
- OAuth authentication integration and K8s dynamic deployment

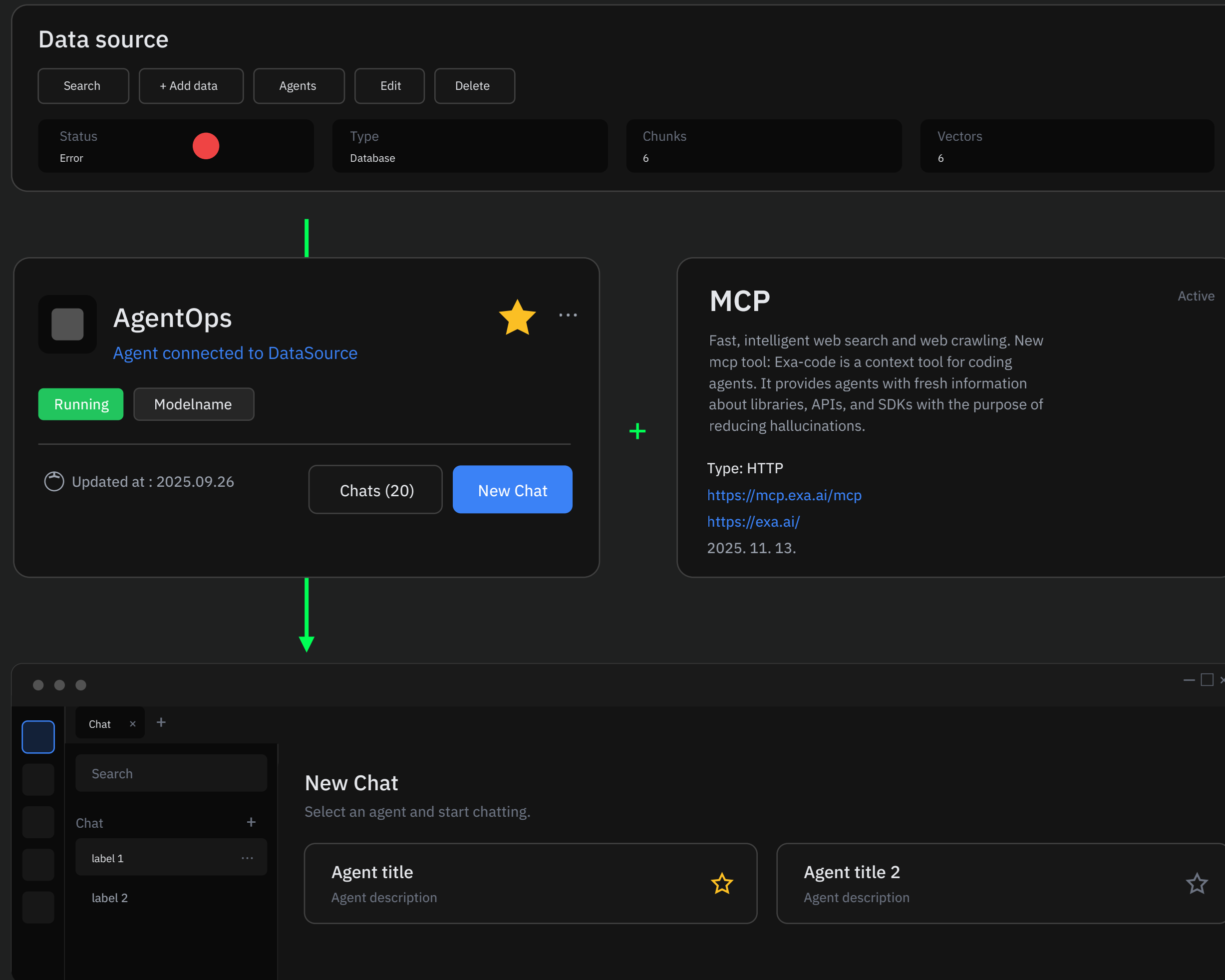
4.

Praxis

Product Composition

Praxis: Executable AI agent in on-premise environments.

Core RAG engine of an executable AI agent that performs decision-making based on internal enterprise data.



Enterprise RAG Engine

- An enterprise RAG engine that automatically collects and analyzes corporate documents, policies, and DBs.
- Providing vector search optimized for internal data and secure responses.

Domain-specific SFT

- Enterprise-customized LLM that learned internal organization terms, policies, and decision-making patterns.
- Improved accuracy via fine-tuning based on meeting minutes, customer logs, and R&D data.

Agent Builder

- Orchestrator that designs AI agents centered on business intent, such as Low-Code/No-Code workflows, events, and Plan & Act decision-making.

Tool Hub & Safety/Audit

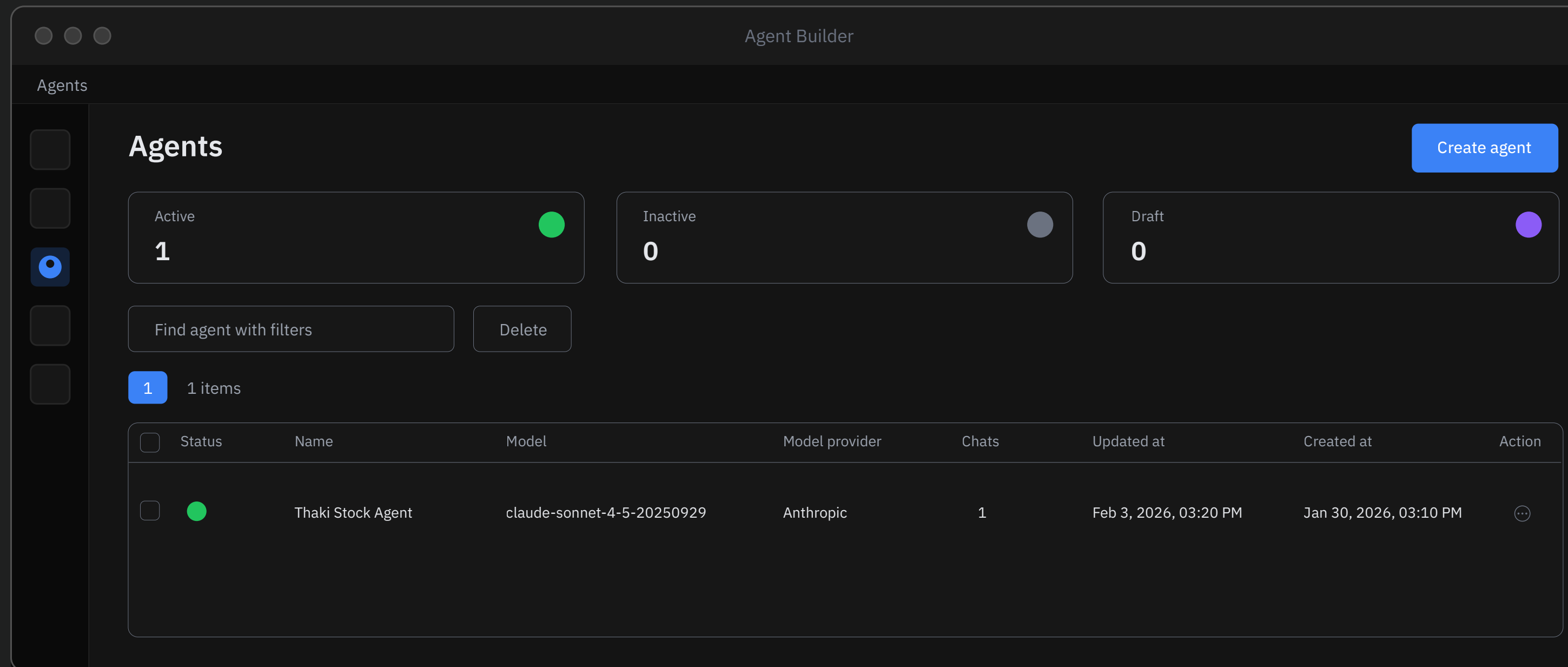
- ERP/CRM/ITSM connection + policy-based filtering.
- PII protection and full action traceability.

Multi-agent Collaboration

- Engine where multiple agents collaborate to solve complex workflows
- Supporting sequential execution, parallel execution, and sub-task decomposition.

Agent Builder

No-Code/Low-Code builder to easily design and manage AI Agents.



LangGraph-based Think → Act → Observe cycle

- Agent execution loop that thinks, acts, and observes results autonomously.

System prompt customization

- Freely set Agent personality and instructions to fit the domain/task.

Data source/MCP tool connection

- Connect knowledge for the Agent to reference and tools for it to execute.

Activation/Deactivation status management

- Deployment control to immediately turn operating Agents on and off.

Enterprise RAG Engine

Hybrid RAG engine that accurately searches internal enterprise data.

The screenshot shows the 'Enterprise RAG Engine' interface for managing 'Datasources'. At the top, there are five status summary cards: 'Completed' (4), 'Error' (0), 'Processing' (0), 'Pending' (0), and 'Draft' (0). Below these is a search bar 'Find data sources with filters' and a 'Delete' button. A table lists 4 items, all with a 'Completed' status (green dot). The table columns are: Status, Name, Type, Documents, Size, Created at, and Action.

Status	Name	Type	Documents	Size	Created at	Action
Completed	Postgre Datasource	Postgres	0/0	-	1/30/2026, 2:55:18 PM	⋮
Completed	GCS Datasource	Google_storage	4/5	-	1/30/2026, 2:54:10 PM	⋮
Completed	File structure Datasource	File	5/5	-	1/30/2026, 2:53:06 PM	⋮
Completed	File semantic Datasource	File	5/5	-	1/30/2026, 2:52:38 PM	⋮

Data source types: FILE, GCS, PostgreSQL

- Integration of various sources such as local files, cloud storage, and relational DBs.

Document processing: Multi-chunking strategy

- Parses even scanned documents and divides them into optimal units.

Hybrid search: BM25 + Dense Embedding + RRF

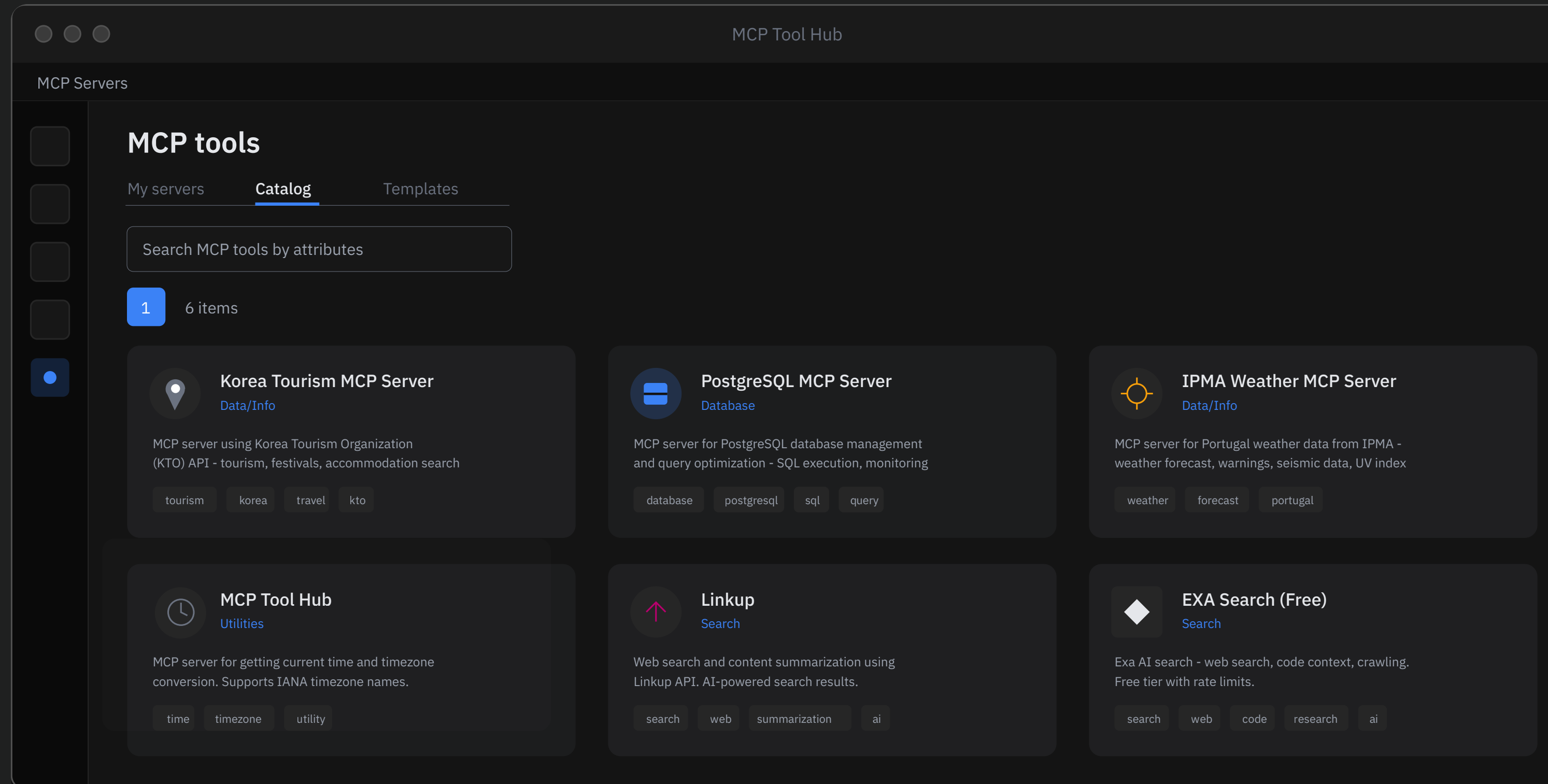
- Maximizes accuracy by integrating keyword and semantic search.

Airflow-based asynchronous pipeline

- Scalable workflow that processes large-scale documents in parallel.

MCP Tool Hub

A tool hub that expands the execution power of the Agent by linking with external business systems.



Server types:
HTTP (external), STUDIO (Container)

- Can link both external APIs or tools inside containers.

Template-based fast server creation

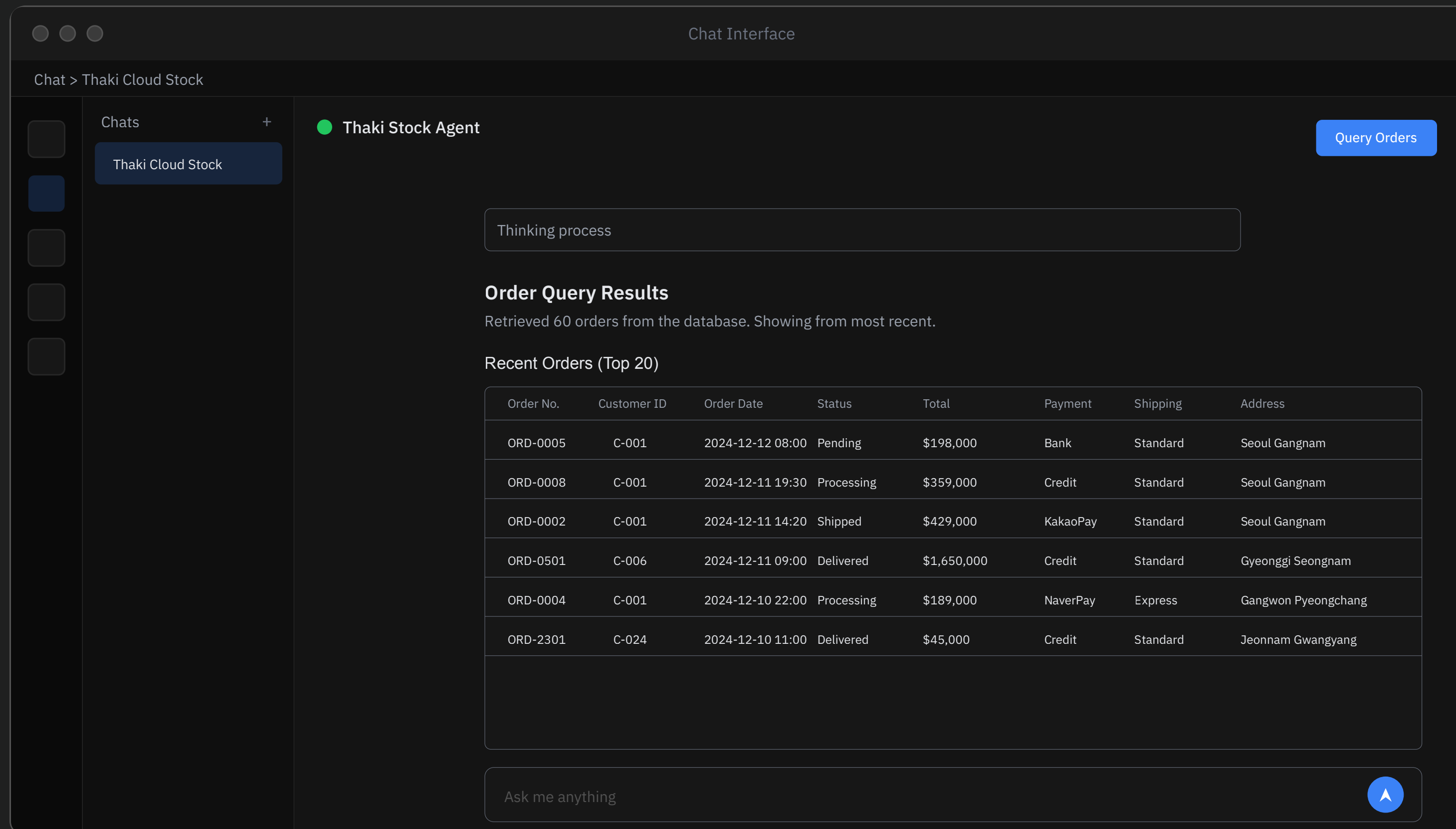
- Deploy MCP servers within minutes with pre-defined templates.

OAuth authentication integration, execution logs

- Security authentication and tracking of all tool call records.

Chat Interface

An intuitive interface to process work while conversing with the Agent in real-time.



WebSocket real-time streaming

Session management and history

- Maintain conversation context and view previous conversation history.

RAG source citation display

- Clearly display the source of the original document that is the basis for the answer.

UI Blocks (charts, tables, code, etc.)

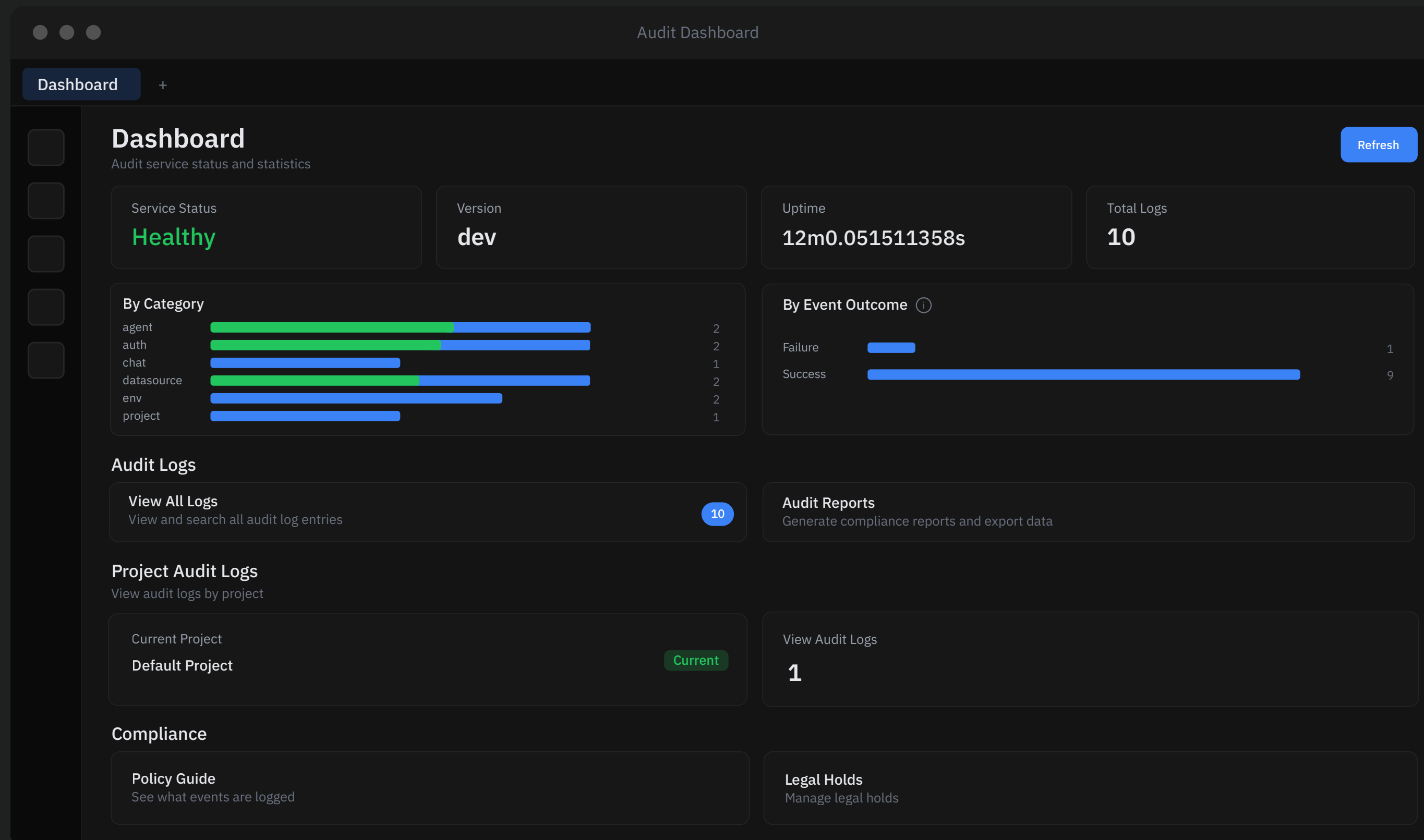
- Render rich responses with visual elements besides text.

5.

Security · Operations

Provides a reliable audit trail by recording all activities immutably.

Thaki AI Agents Audit Service is a next-generation audit log system capable of high-performance large-capacity processing while meeting SOC 2 Type II compliance.



Hash Chain integrity guarantee

- Each log is connected in a chain including the SHA-256 hash of the previous log.
- If tampered with, the chain breaks and is immediately detected.

Full multi-tenancy isolation

- Complete separation of data between tenants with RLS (Row Level Security)
- Applying independent retention policies per tenant.

Legal Hold

- Protects legal evidence with indefinite log retention and automated deletion-prevention.

High-performance lossless processing

- 10,000+ RPS processing, guarantees data losslessness even during service failures with WAL (Write-Ahead Logging).

6.

Business Impact

Praxis

Praxis provides reduced Agent construction time, improved search accuracy, and SOC 2 level compliance through hybrid RAG, MCP tool integration, and on-premise deployment.

Agent Builder

↑ Efficiency

Enterprise RAG Engine

↑ Accuracy

MCP Tool Hub

↓ Cost

Agent Builder

- Reduction of development resources
- Fast construction of domain-specialized Agents.

Enterprise RAG Engine

- Maximization of data utilization rate
- Improvement of search accuracy.

MCP Tool Hub

- Reduction of existing system linkage costs.

Appendix #1

High-Level Architecture

Appendix #1

High-Level Architecture

Frontend

AI agent self-service portal based on React · TypeScript · Vite/Rspack · TailwindCSS

AI Agent Builder · Conversational Chat (WebSocket) · Data source mangement · MCP server management · Project/IAM

Service Layer

FastAPI · Python · LangGraph · Pydantic v2 · Go(Audit Service)

Agent Orchestration · RAG Pipeline · Chat Streaming · MCP server lifecycle · Auth/ABAC · Multi-tenancy

Orchestration & Pipeline

Apache Airflow(KubernetesExecutor) · Kubernetes · Helm · Argo CD

Document processing pipeline(Parsing → Chunking → Embedding → Vector storage) · MCP server K8s deployment · Multi-cluster management · GCS/RDB synchronization

Workload Management

PostgreSQL · Qdrant(Vector DB) · NFS · Anthropic Claude · vLLM · Jina Embeddings

Relational data · Vector embedding & hybrid search(BM25+Dense) · File Storage · LLM inference · embedding