Getting Started with Agentic AI and MCP servers

Natale Vinto

Red Hat

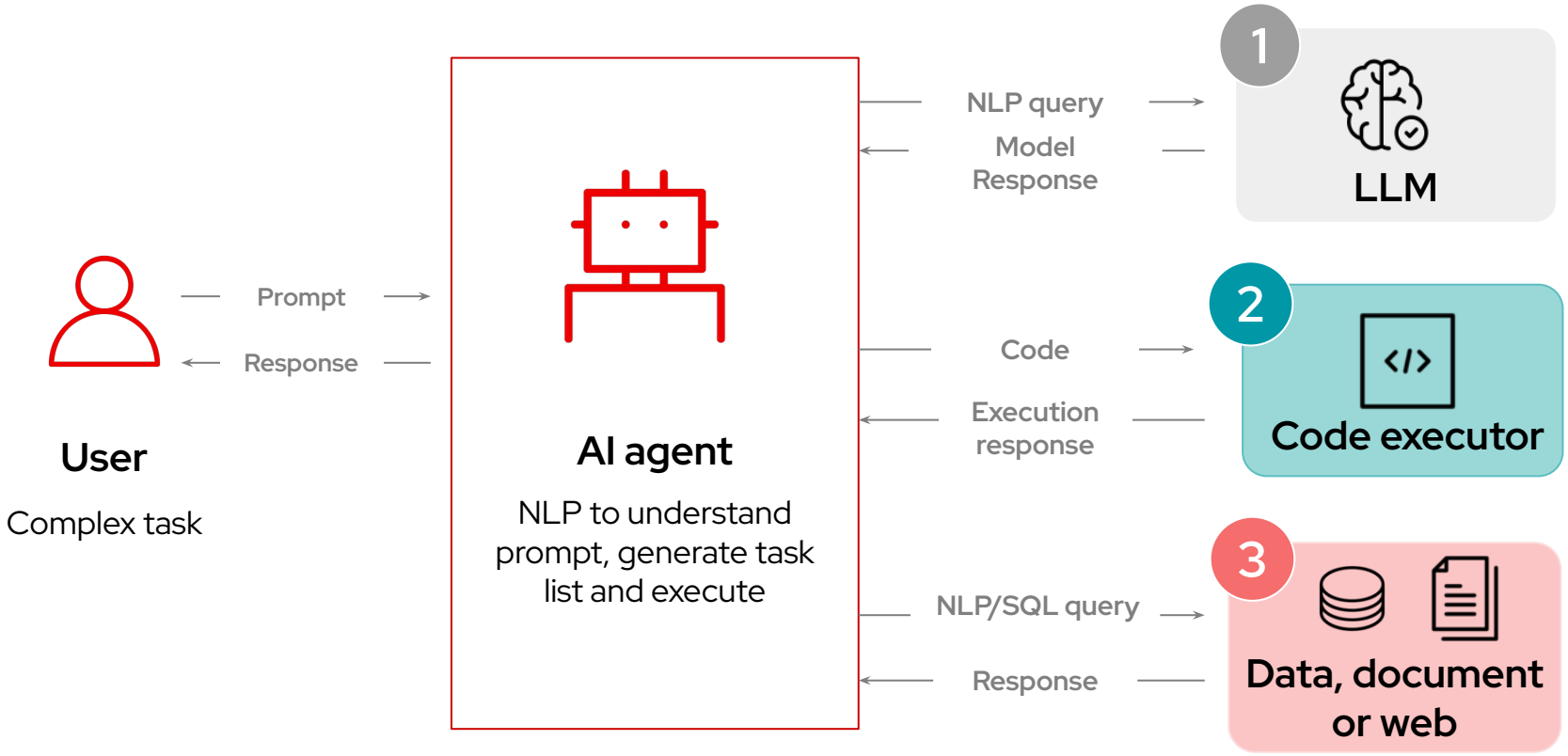
Codrin Bucur

Red Hat

Intro to Agentic AI

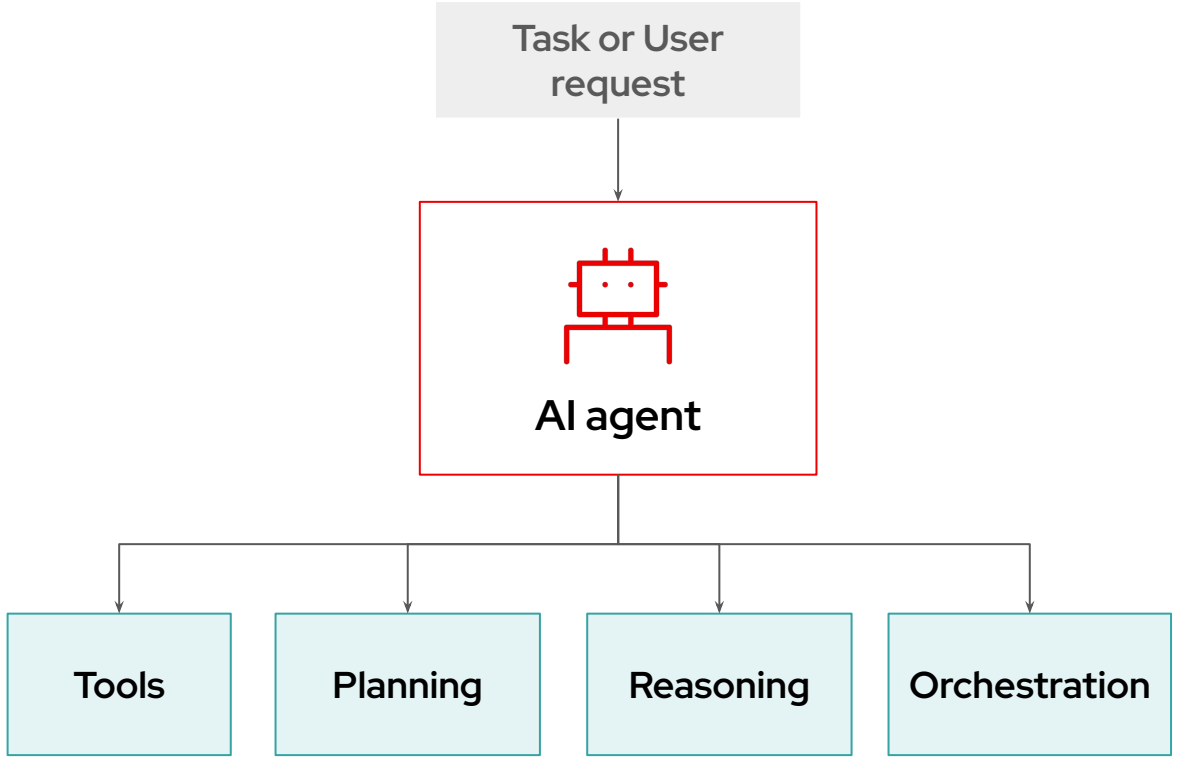
AI agents integrate models, functions & tools

Gen AI Models, Predictive AI Models, Code Functions, Search & more



Red Hat

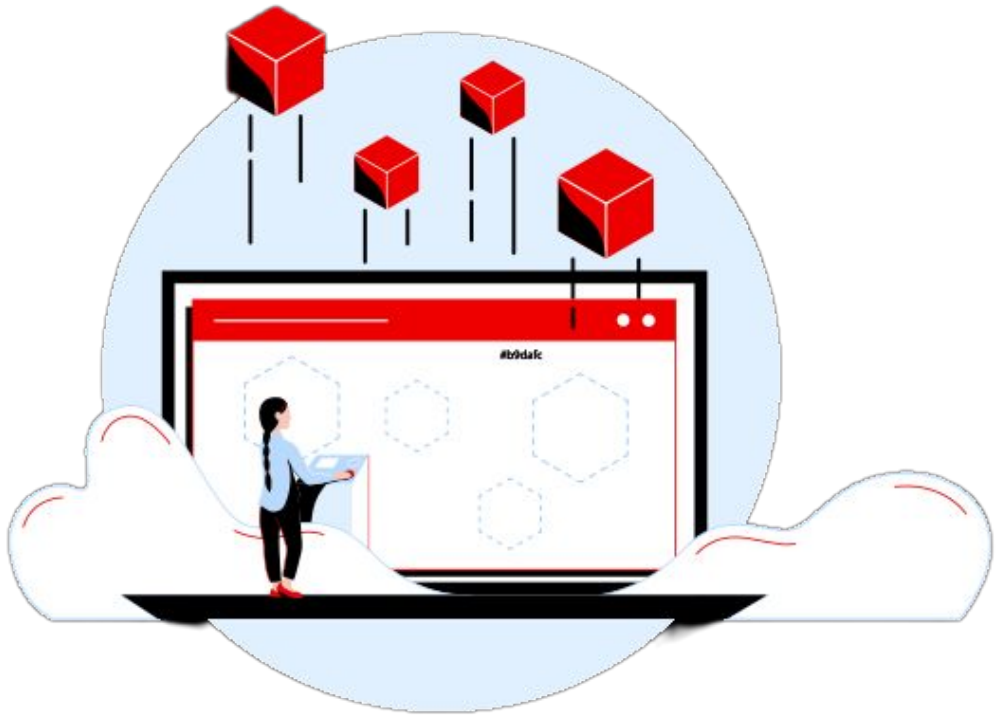
The components of an AI Agent system



- **Tool Utilization:** Leverages external tools to gather data and perform tasks.
- **Planning and Execution:** Develops and executes multistep plans to achieve goals autonomously.
- **Reasoning:** Applies logic and contextual understanding to make informed decisions.
- **Orchestration:** Coordinates actions, tools, and agents to dynamically adjust and complete tasks.
- **Communication protocols:** enables the connections between the components.

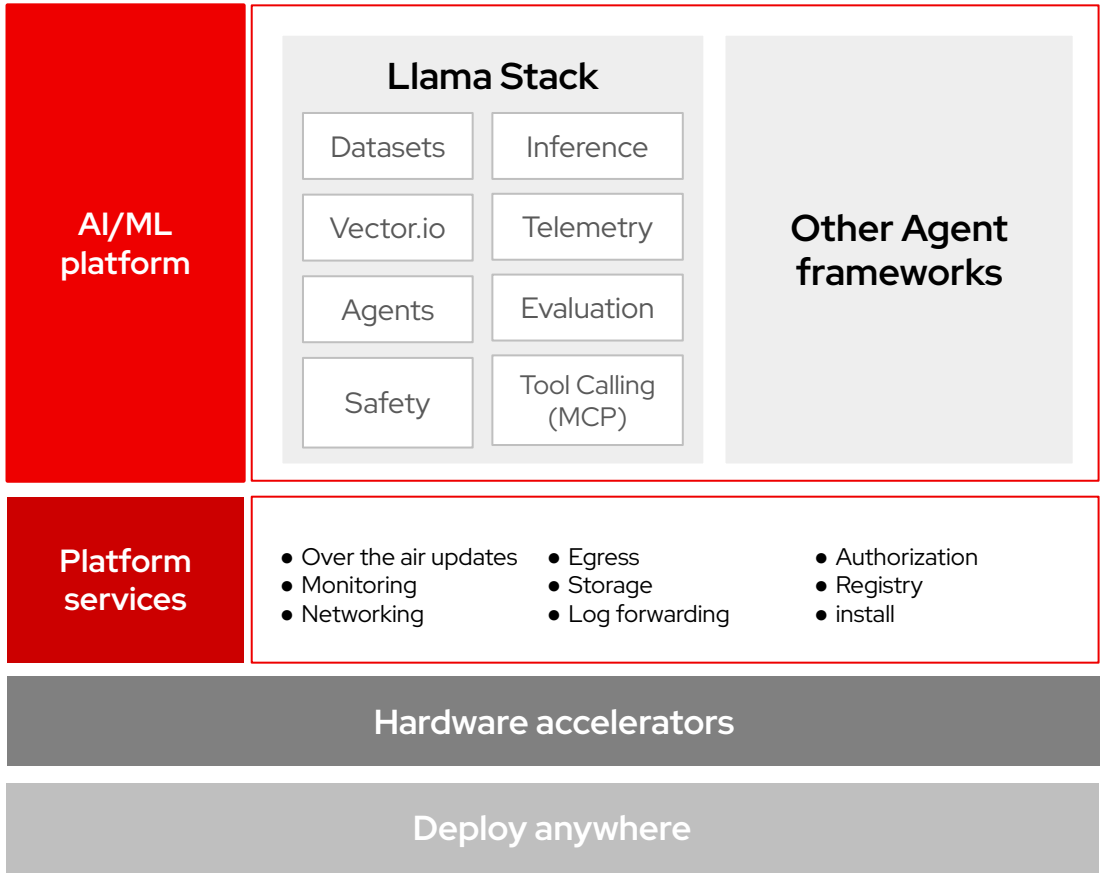
Red Hat AI provides an agile, stable foundation to accelerate the development and deployment of AI agentic workflows.

- Offers built-in agent frameworks with Llama Stack, and standardized communication protocols (MCP).
- Provides the flexibility to integrate preferred tools like LangChain and Crew AI.
- Allows running and managing agents as microservices.
- Simplifies production deployment by managing LLM serving and scaling.



Red Hat

A modular approach to building AI agents

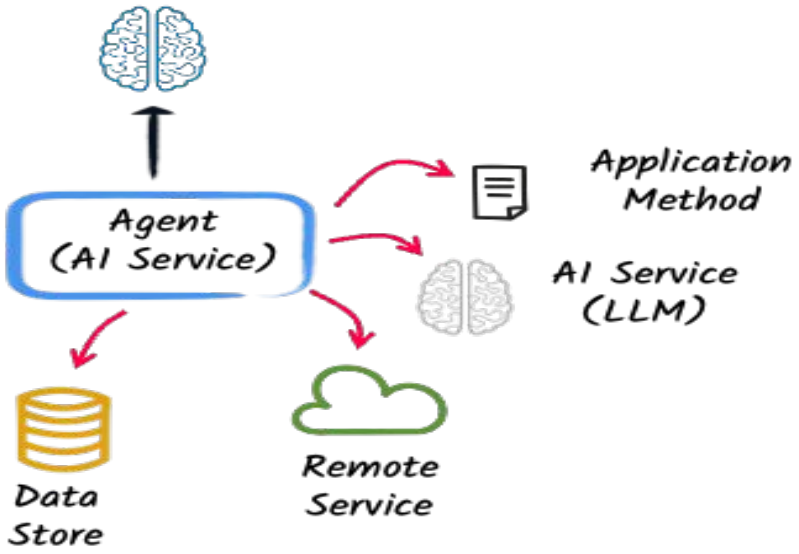


Red Hat AI allows to:

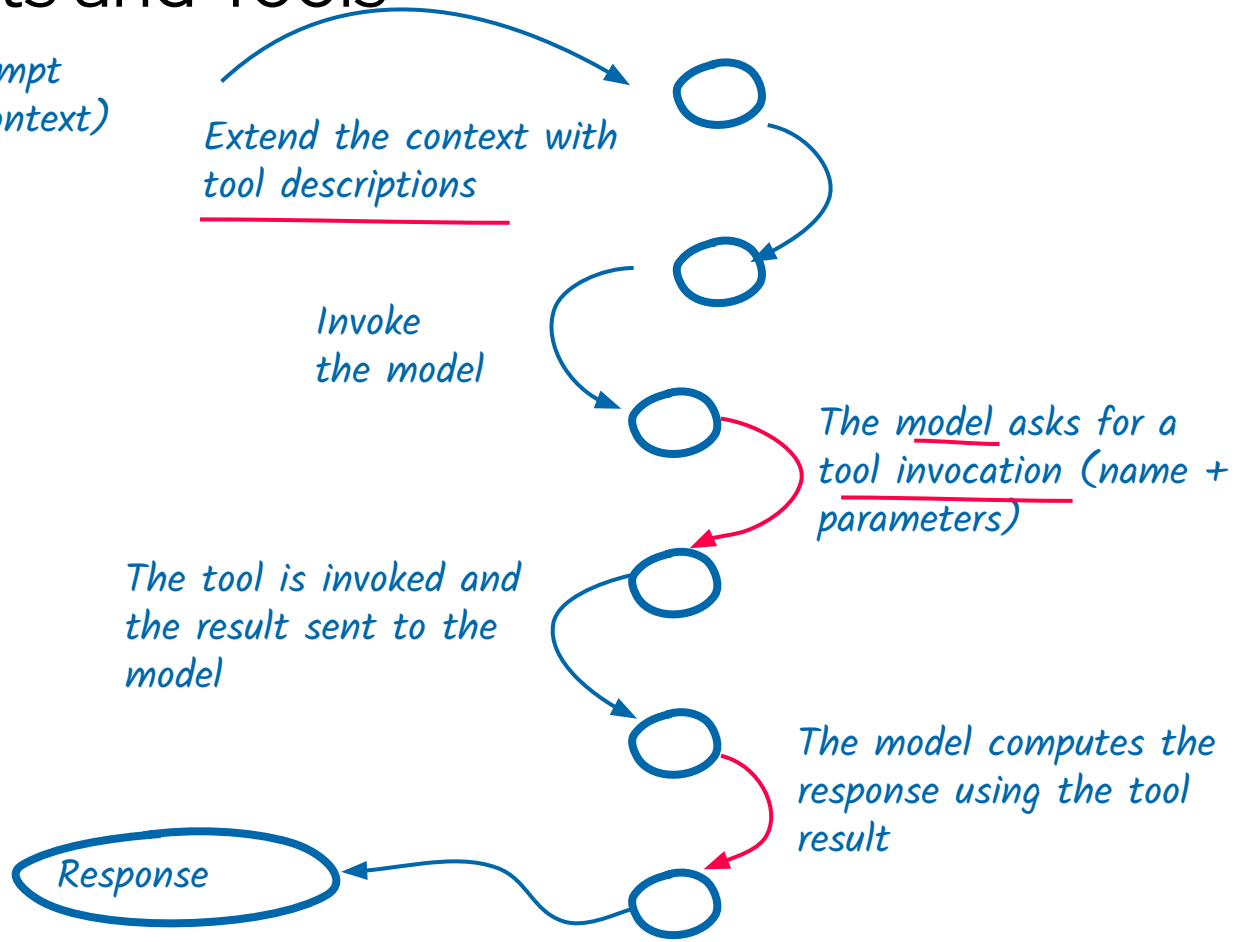
- Build agents using **Llama Stack's native capabilities and implementations**.
- **Bring compatible Llama Stack implementations** to OpenShift AI.
- **Use your own agent framework** and selectively incorporate Llama Stack APIs.
- **Build with Core Primitives** and manage your own agent framework as a standard workloads.

Red Hat

AI Agents and Tools

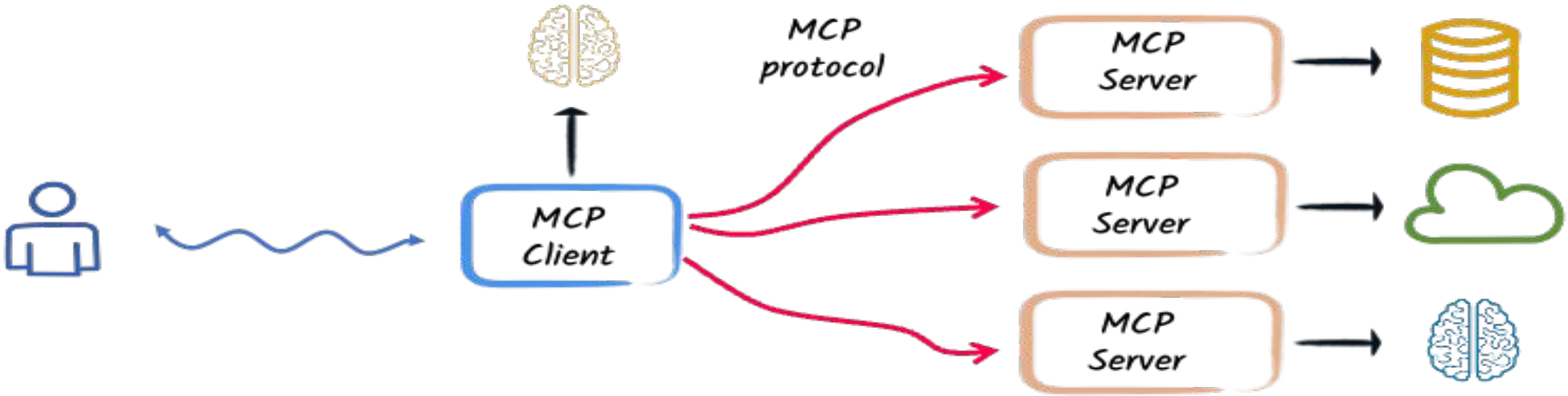


Tools require memory and a reasoning model



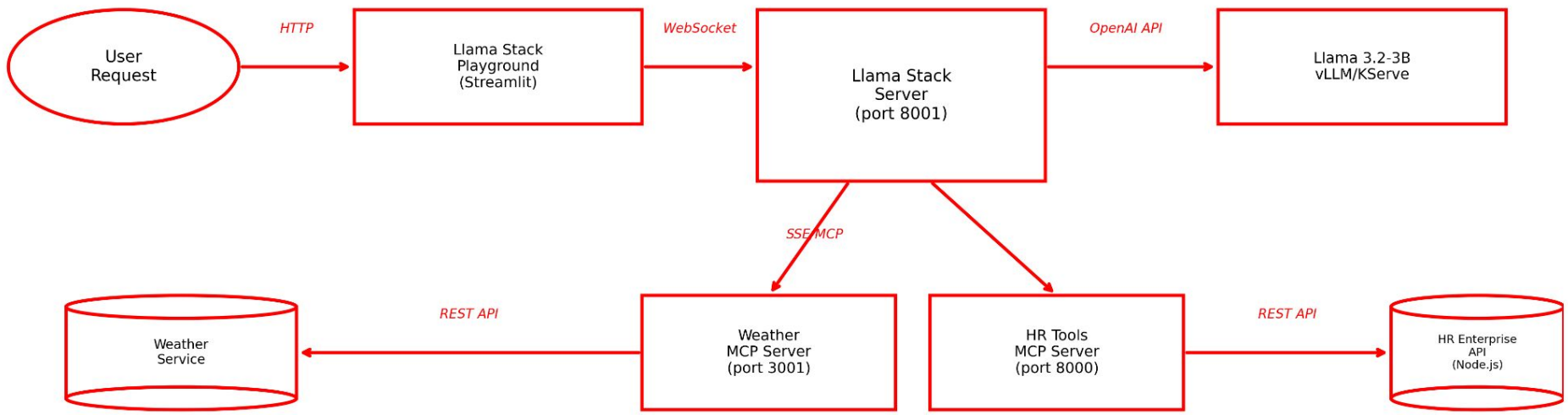
Model Context Protocol (MCP)

The Model Context Protocol,introduced by Anthropic in November 2024, is an **open standard** that enables developers to build secure, two-way connections between their data sources and AI-powered tools.



Instead of exposing tools from your code,
discover and use remote services

Llama Stack with MCP Servers Architecture



OpenShift / Kubernetes Infrastructure
Pods • Services • Routes • InferenceService • ServingRuntime

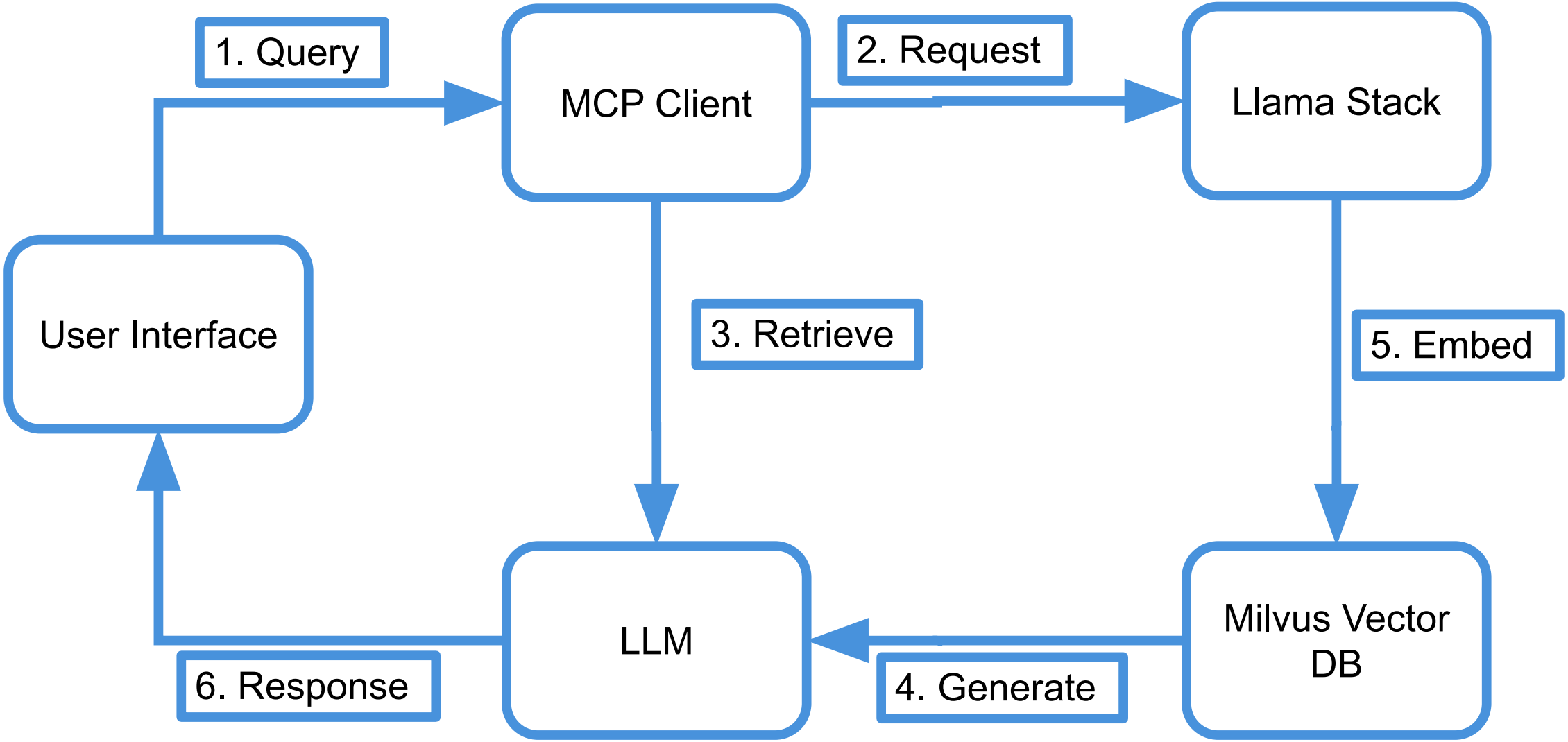
Key Technologies:

- Model Context Protocol (MCP): Tool integration framework
- Server-Sent Events (SSE): Real-time communication
- vLLM: High-performance LLM inference engine
- Helm Charts: Kubernetes deployment automation

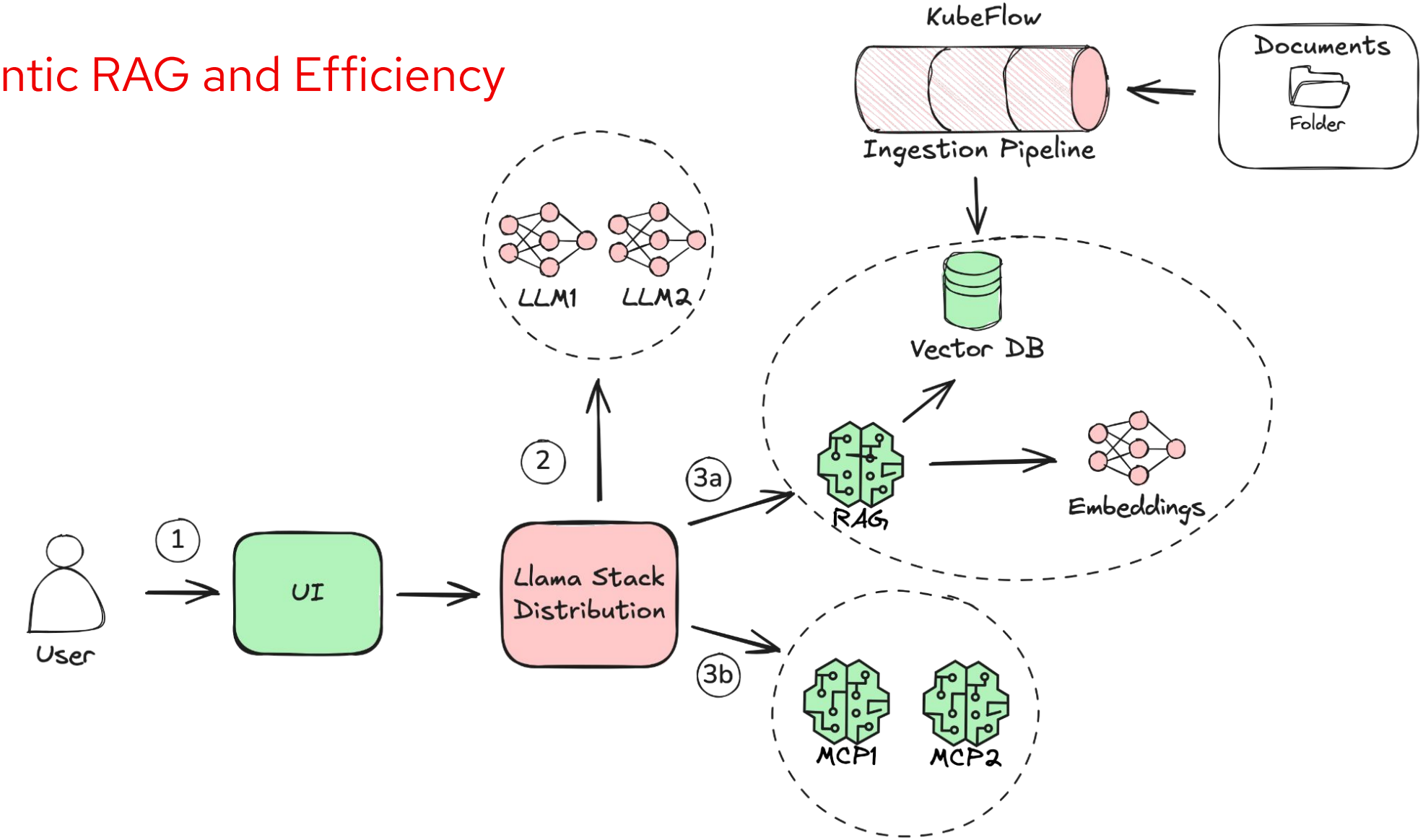
Available Tools:

- Weather forecasts and alerts
- Employee vacation balance queries
- Vacation request creation
- HR data access and management

Agentic AI Lab Architecture

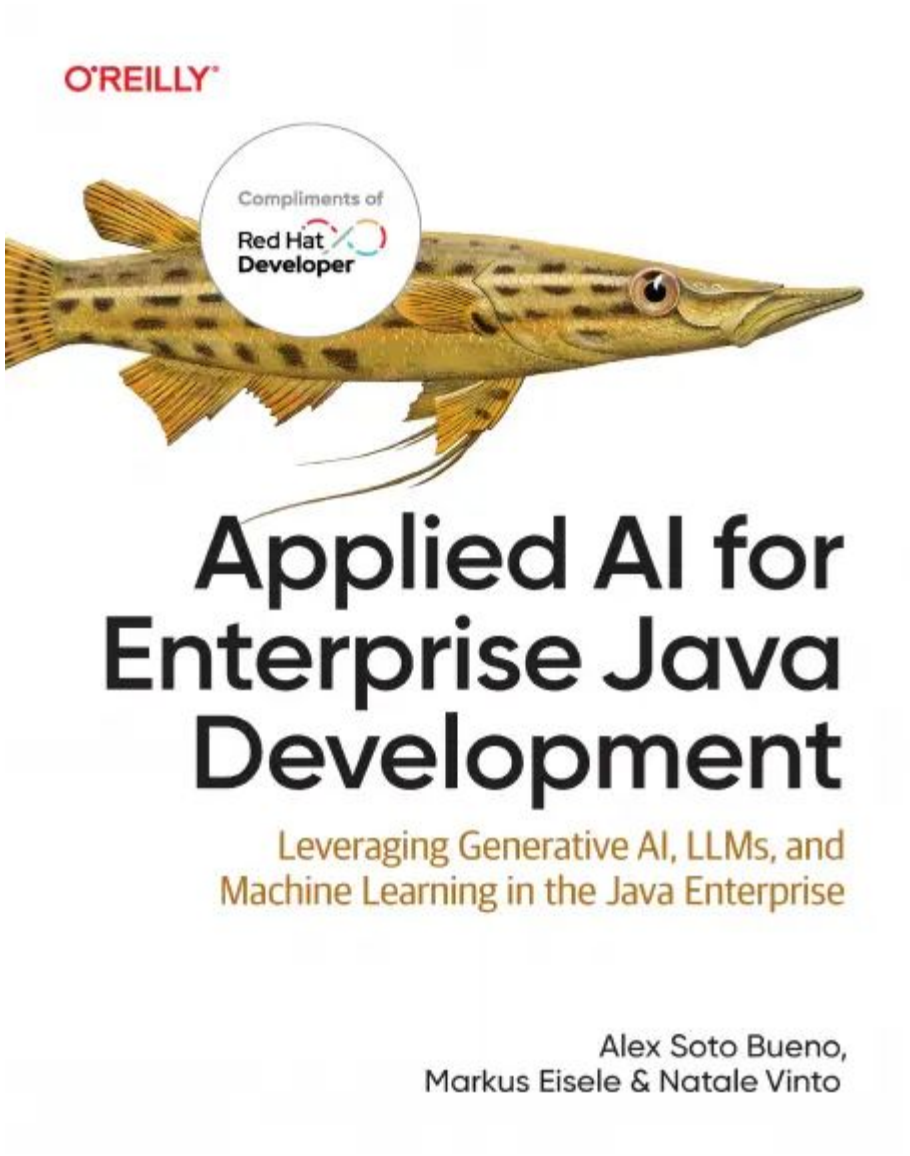


Agentic RAG and Efficiency



red.ht/agentic-lab-zurich

Password: **redhat**



Download the free e-book from:
developers.redhat.com

Red Hat

red.ht/rhsc-session-6

Session 6: 16:00 - 17:00

Your feedback is important!

Scan the QR-code, select the session and evaluate the presentation. **Thank you!**