Backstage CON

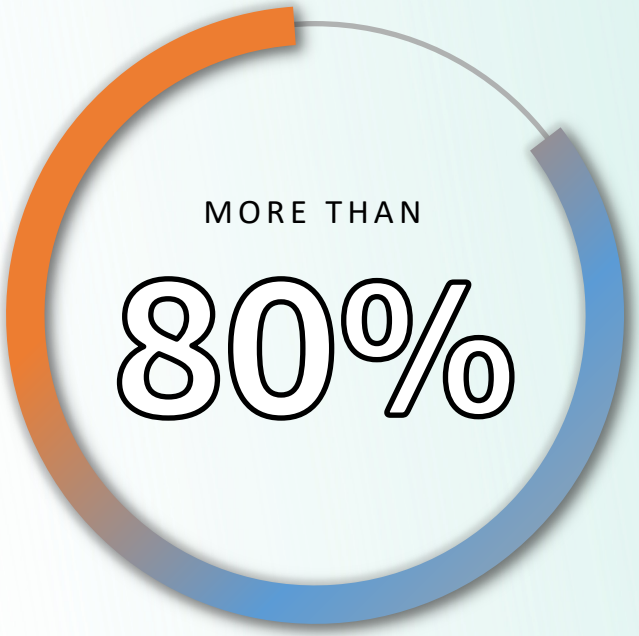
NORTH AMERICA

Backstage CON
NORTH AMERICA

Empowering LLMs with Backstage:

Broader Insights Driven by the Developer Portal

Niall Thomson, AWS



ACCORDING TO GARTNER, INC.®

# of enterprises will have used generative AI APIs or deployed generative AI–enabled apps by 2026[1]

[1] Gartner, “More than 80% of Enterprises,” October 11, 2023.

How does a foundational model work?

Unlabeled data

Foundation model

- Text generation
- Text summarization
- Information extraction
- Image generation
- Chat bot

Broad range of general tasks

Platform engineering use-cases

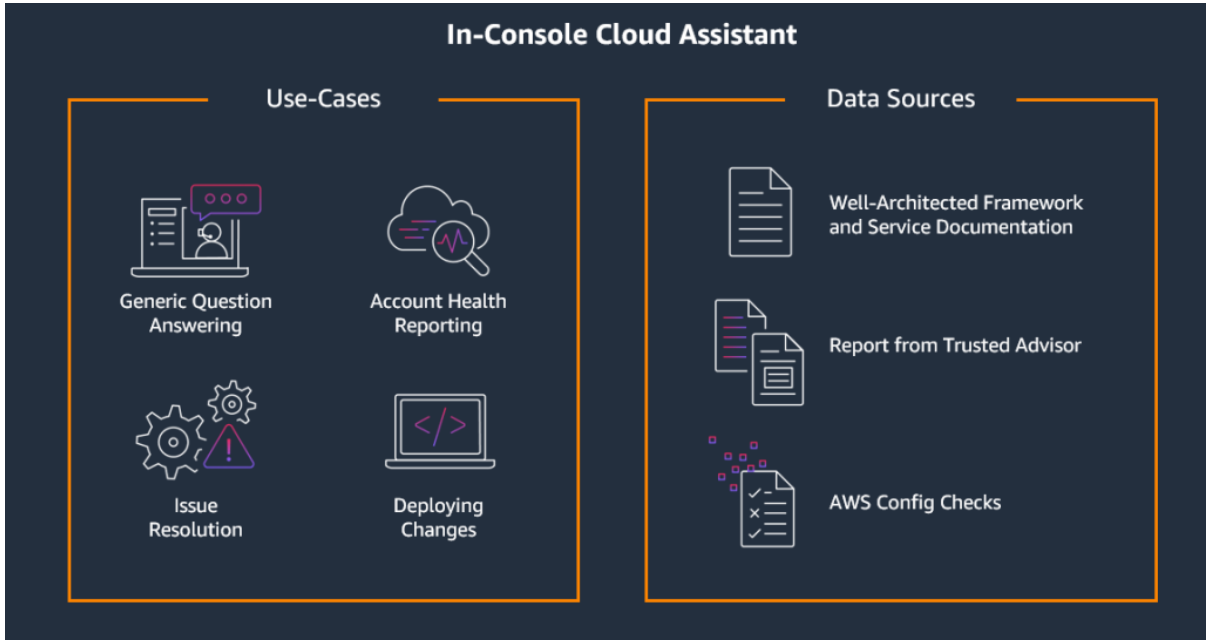
Enablement

Provide targeted education to developers based on their needs

Issue resolution

Enable developers and platform engineers to diagnose and remediate issues

BMW Generative AI Assistant



Why Backstage?

Contextualize knowledge

Rich source of additional data

Platform integrations

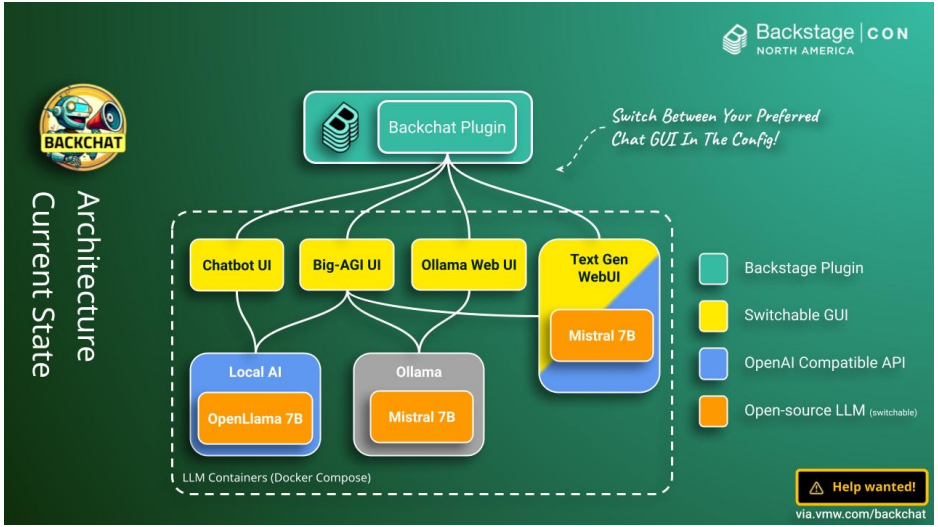
Diverse sources of data to augment responses

Developer workflows

Leverage investments in existing portal to surface capabilities

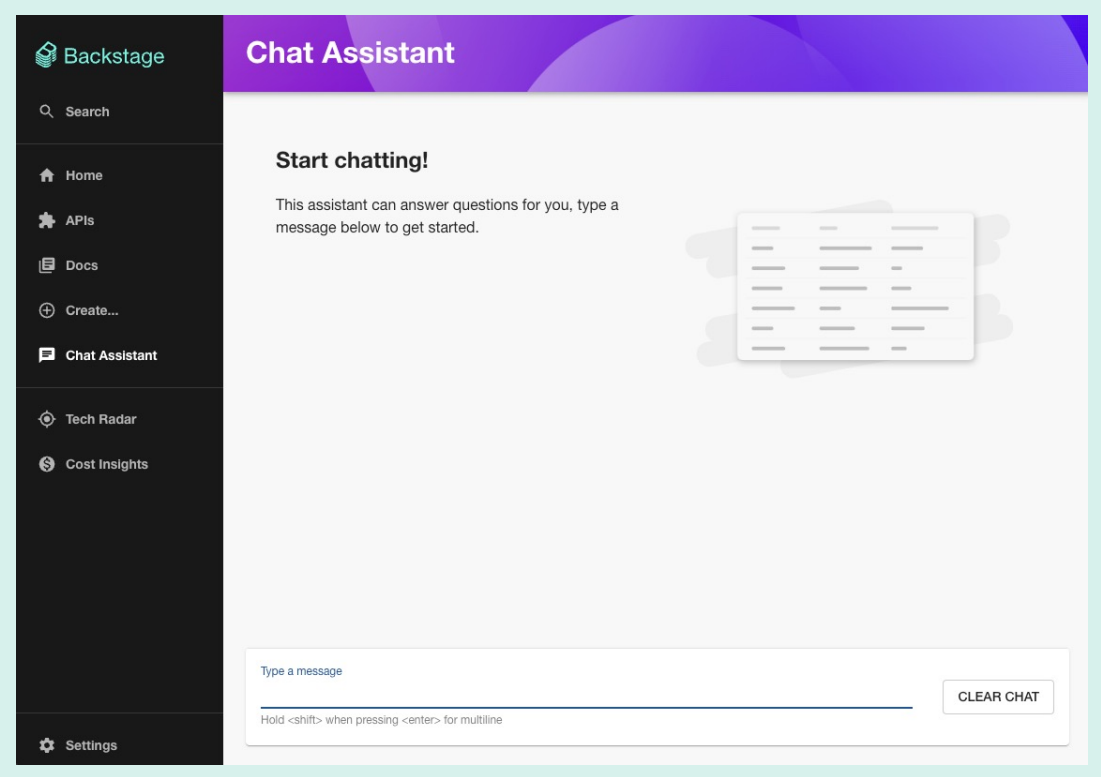
Backchat

- Embed chat GUI in Backstage
- Leverage existing GenAI projects
- Quickly experiment



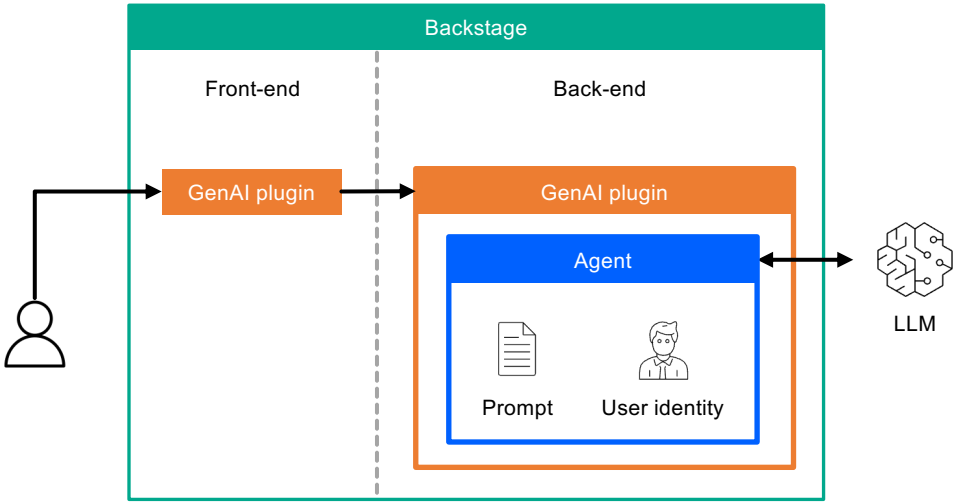
https://github.com/benwilcock/backstage-plugins-backchat

Let’s build!


Backstage
Search
Home
APIs
Docs
Create...
Chat Assistant
Tech Radar
Cost Insights
Settings
Chat Assistant
Start chatting!
This assistant can answer questions for you, type a message below to get started.
Type a message
Hold <shift> when pressing <enter> for multiline
CLEAR CHAT

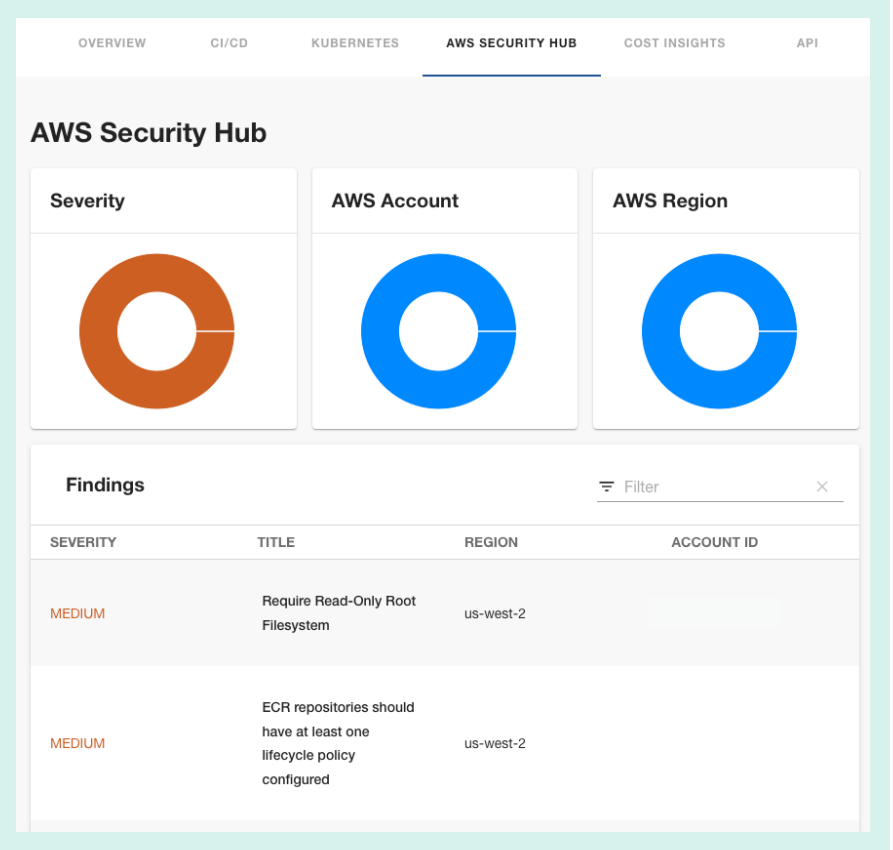
Backstage plugin architecture

- Front-end shows assistant, sends query to back-end
- Back-end is configured with system prompt
- Prompt is sent to LLM, response streamed to browser



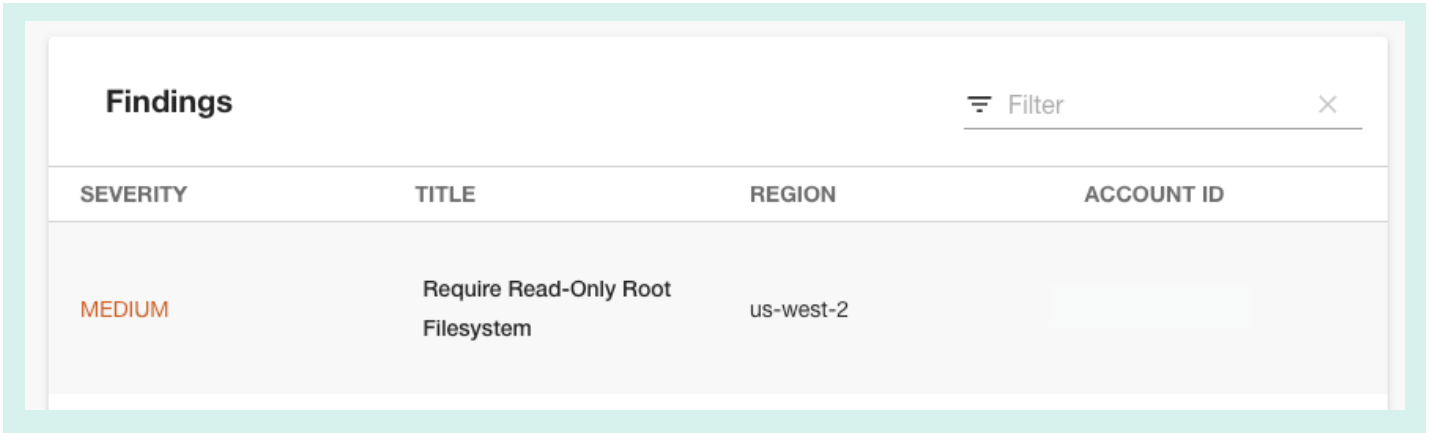
Scenario: Security finding remediation

- Backstage plugin to show findings from AWS Security Hub
- Results are contextual to the workload
- Queries the API directly to get real-time data



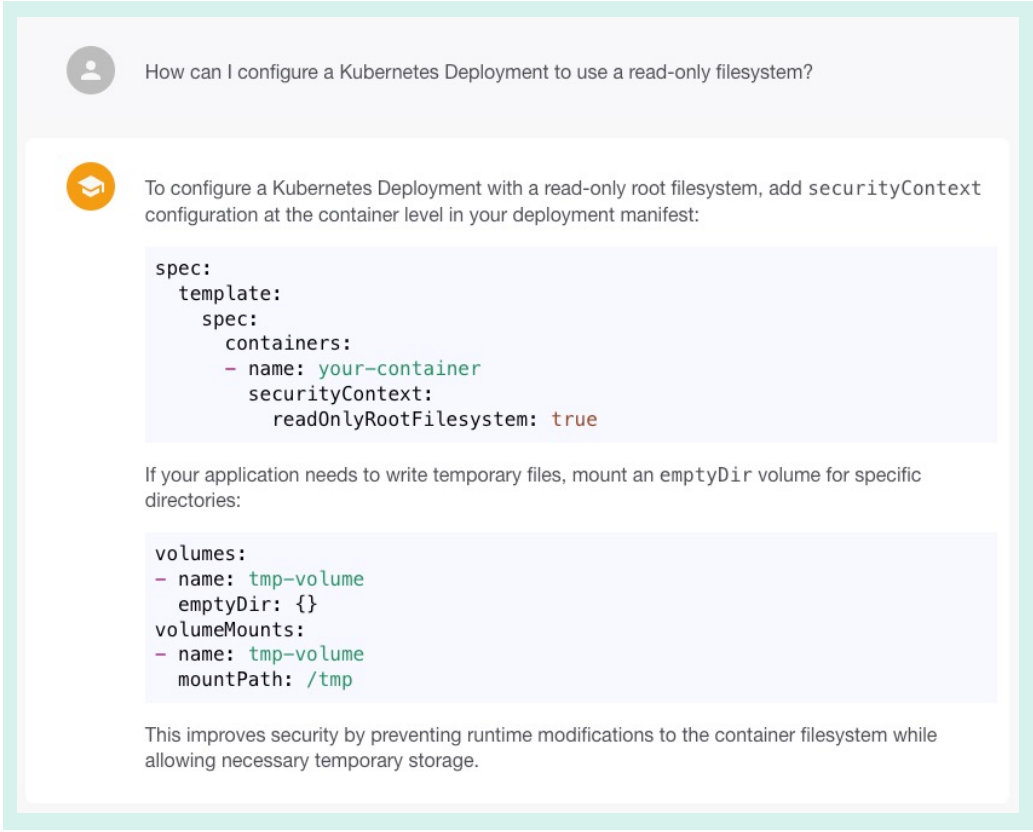
Scenario: Security finding remediation

Task: Resolve this finding for our workload



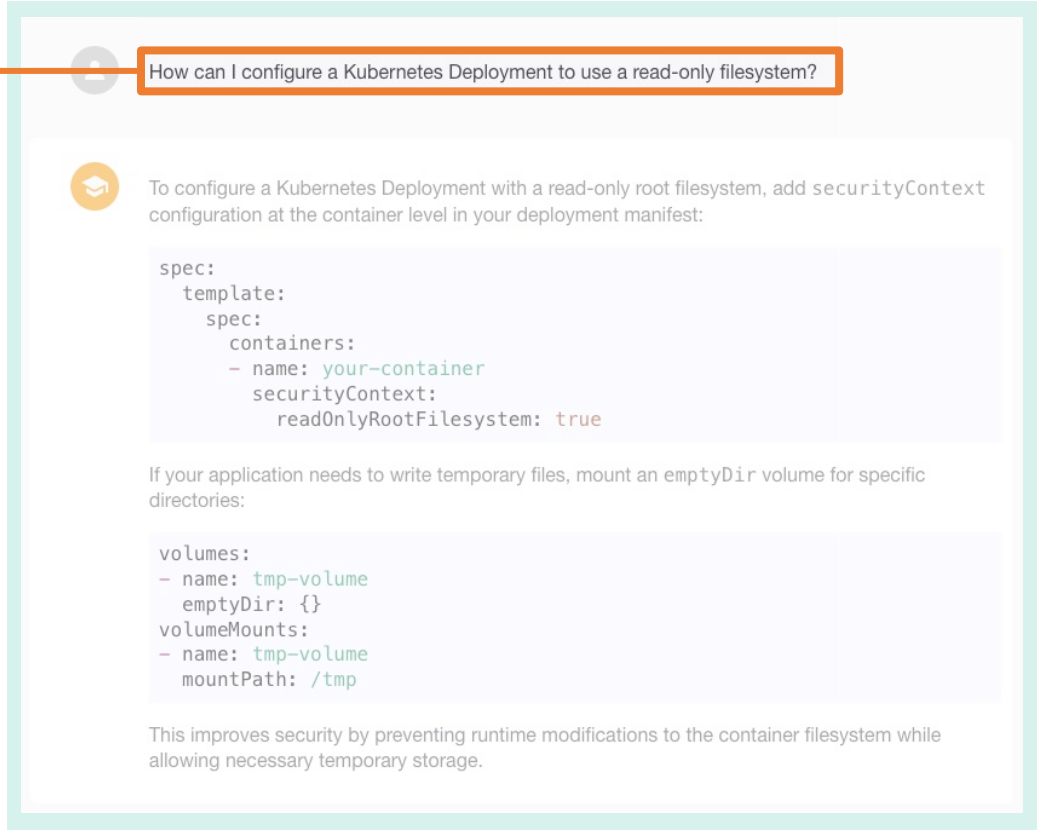
Basic Q&A

- Rely on the LLMs knowledge
- Manage developers LLM interactions
- Custom prompts can provide contextual foundation



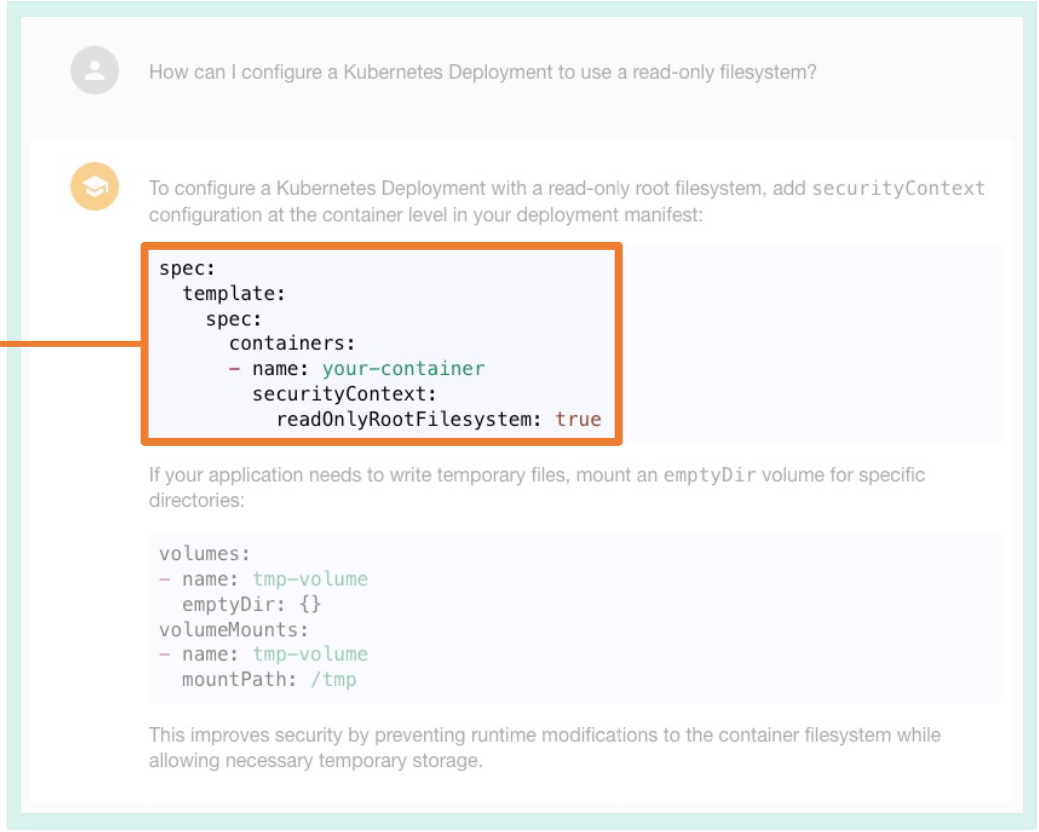
Missing context

How does the user know they need this?

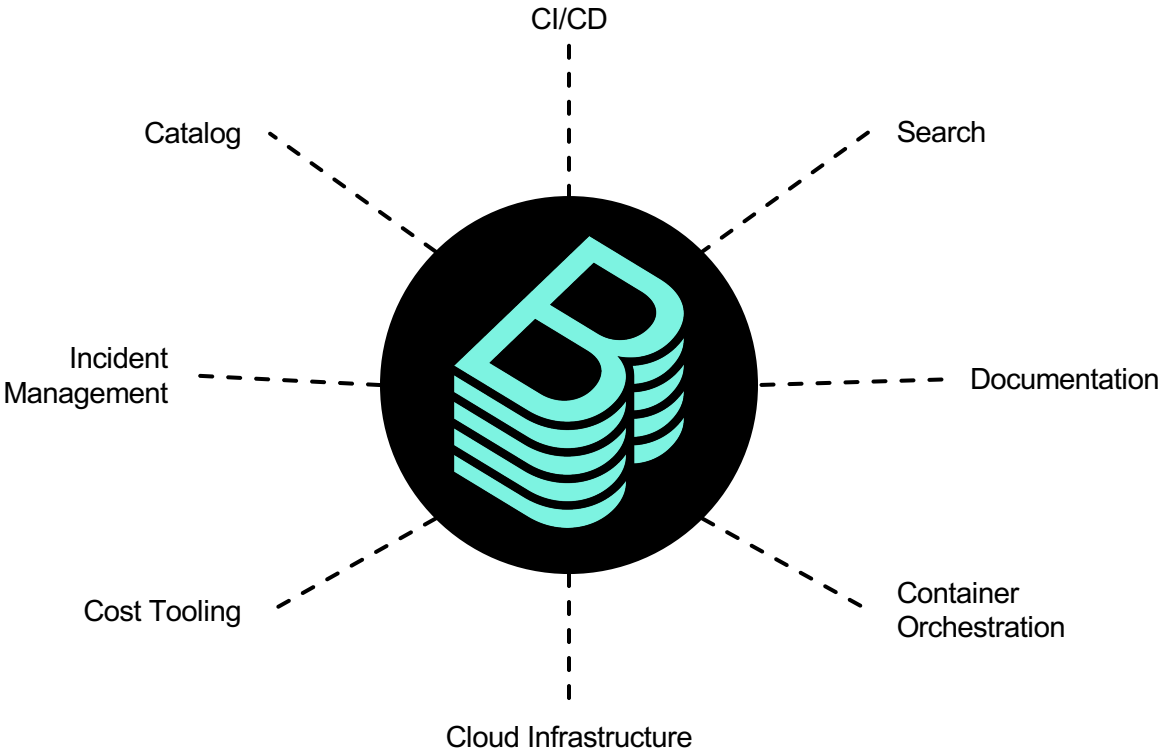


Missing context

Our organization uses custom Helm charts



Backstage integrations

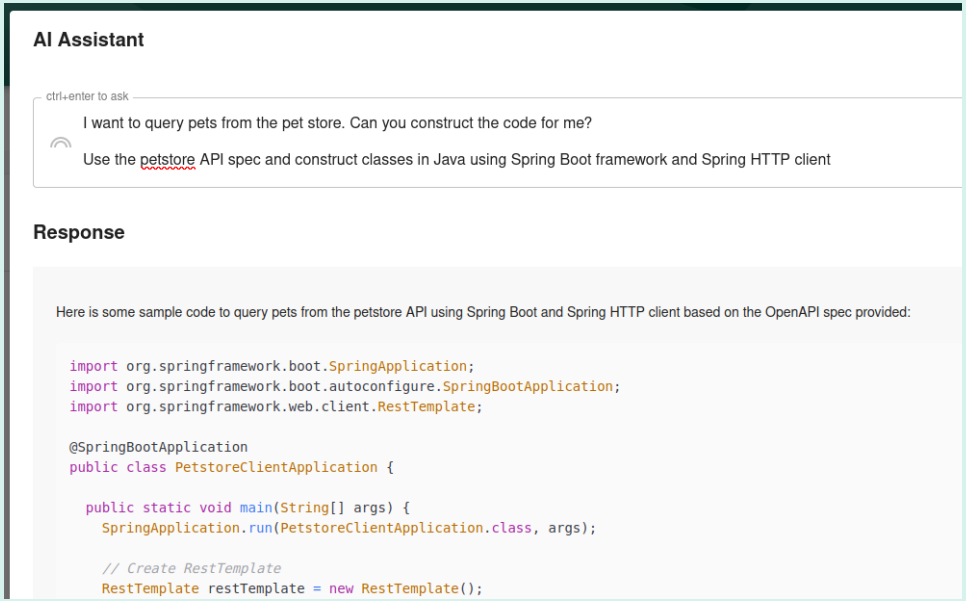


Roadie – RAG AI

First-class Backstage plugin

Contextualize entities, TechDocs and other data

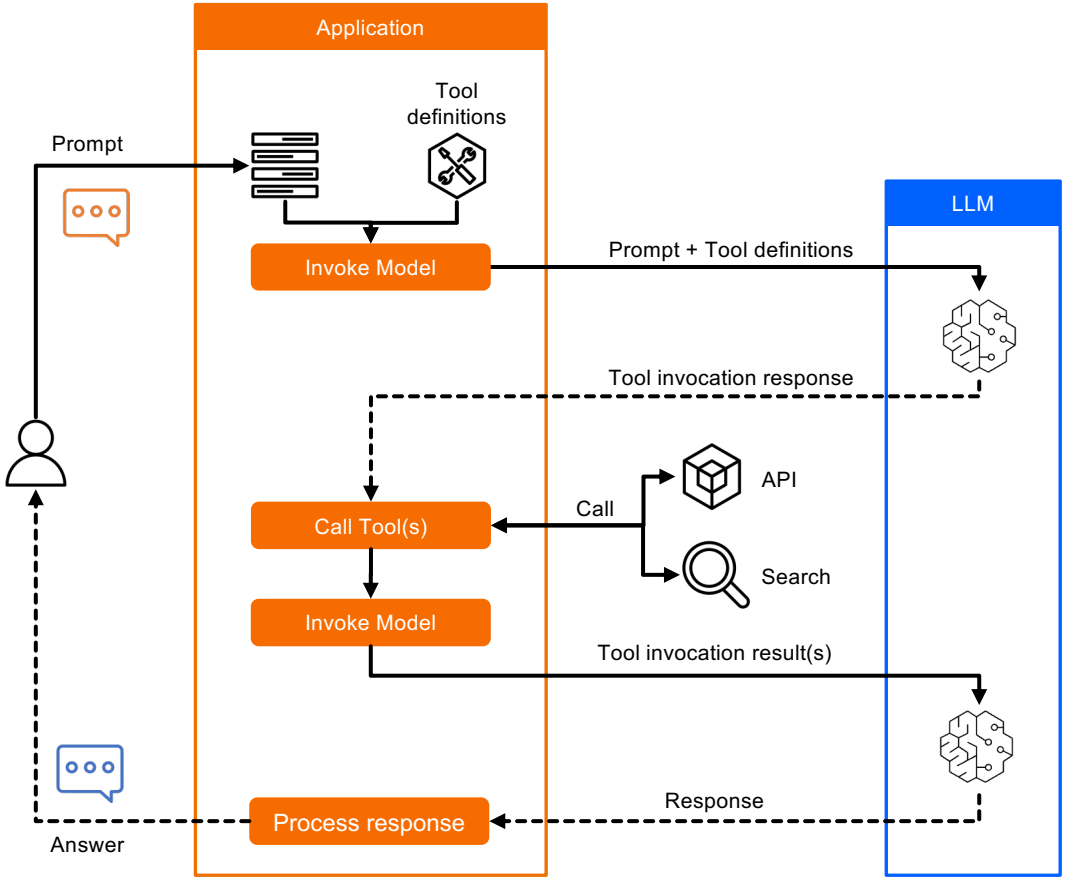
Modular architecture



https://github.com/roadiehq/roadie-backstage-plugins

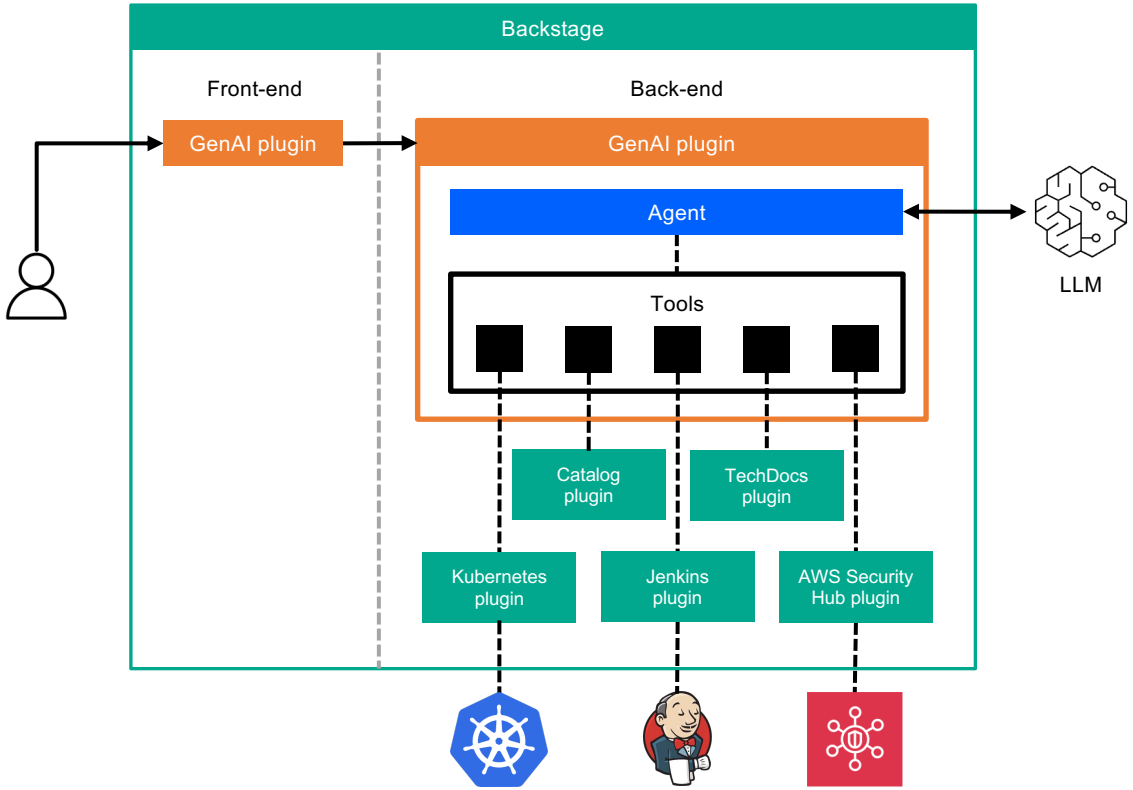
Tools/Function calling

- LLMs retrieve context from external systems
- Model decides what it needs to use
- Usually delegates back to calling application



Tools in Backstage

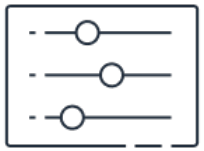
- Back-end plugin introduces tool definitions
- Tools can wrap existing plugins
- Agent sends tool definitions and invokes tools when requested



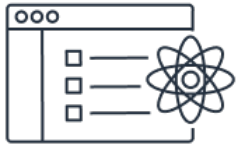
DEMO

Emerging principles

Start simple



Customize for your organization



Guide user adoption

AWS Generative AI plugin on GitHub

- Experimental Generative AI plugin for Backstage available
- Accelerate adopting chat assistant and contextual assistance
- Modular and extensible

https://github.com/awslabs/backstage-plugins-for-aws

Thank you!