SOFTRAMS

Platform Engineering 2024

A Practitioner's Playbook

Building A Serverless GenAI Platform

Public Cloud Edition

About Me

Murali Mallina

- Building teams and software for over 25 years
- CTO, Softrams
- Founder & CEO, Teaching For Good
- Telecom, Supply Chain, EdTech, Healthcare and Civic Services
- Germany, UK, France, India, China, Russia and USA
- LinkedIn: /mkmurali
- Twitter : @mkmurali

About Us

Softrams is one of the fastest growing civic digital services firms, supporting a variety of mission critical workloads for various Federal Agencies.

Hello

- Using GenAI for fun?
- Using GenAI @ Work ?
- Building new GenAI applications?
- Bringing GenAI experiences to existing applications?

Before ChatGPT and After ChatGPT

Generated by ChatGPT4o

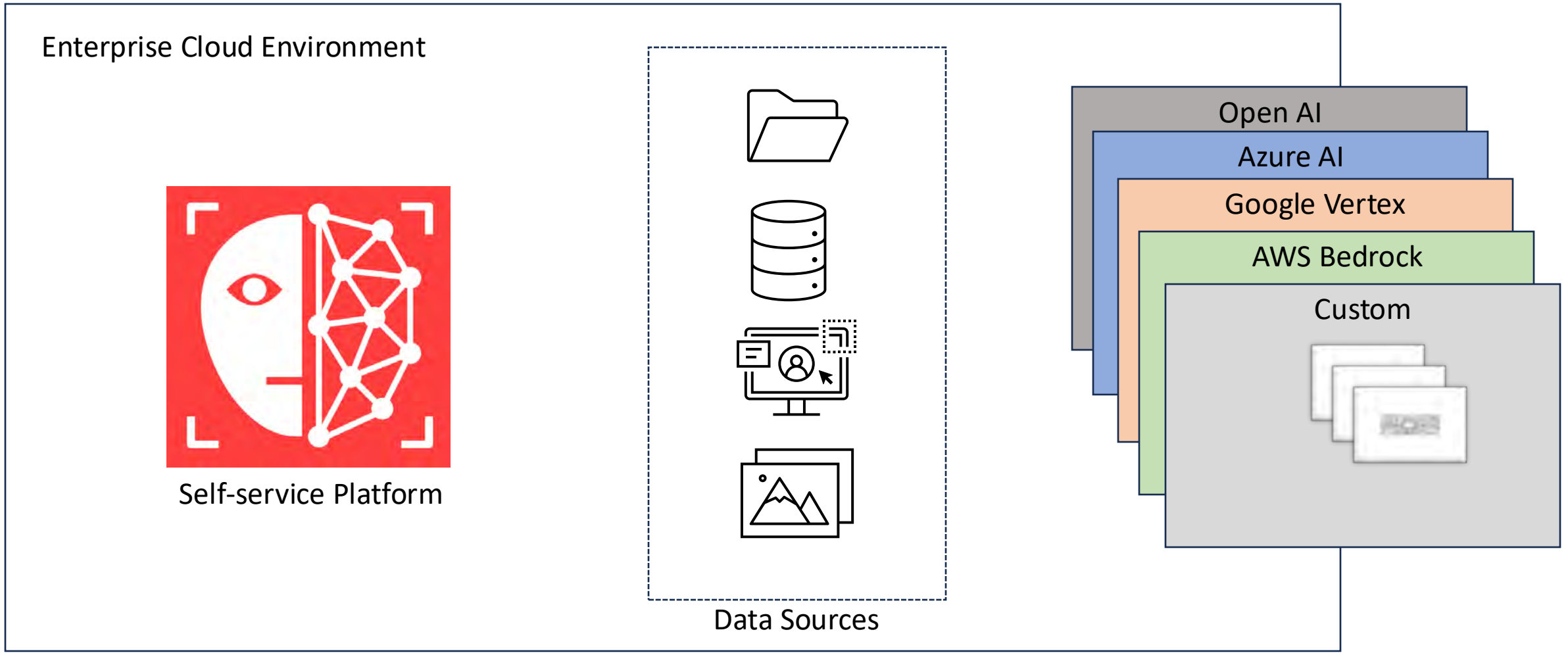
Natural language is the
new programming language

SOFTRAMS

Chat is becoming the
primary interaction model

SOFTRAMS

Serverless GenAI Platform



Works securely with the enterprise data from **within and** leverages Foundational Models and custom models alike.

Generative AI Applications

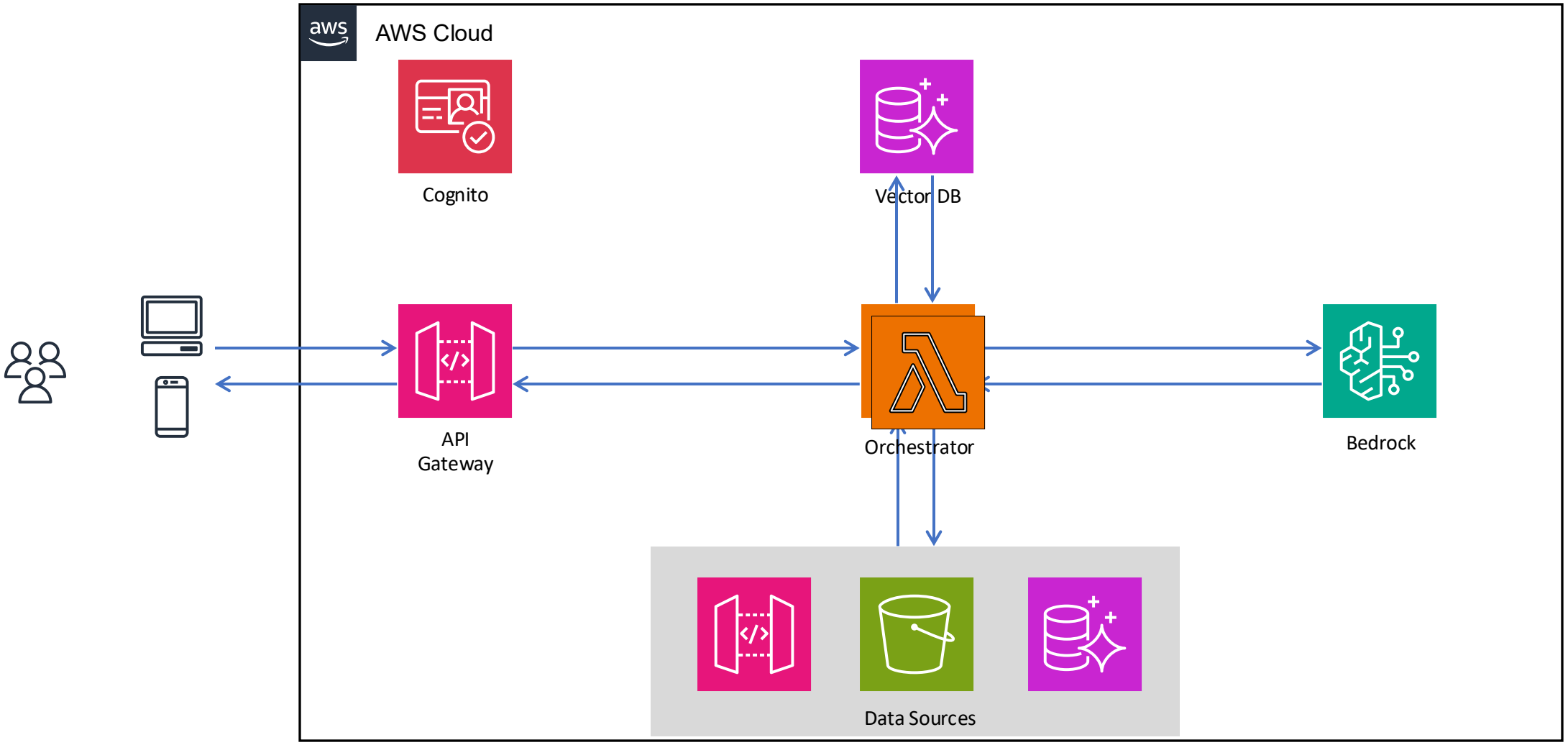
Chatbots
Bots with grounded knowledge

AI for BI
Agents with data access
for Business Intelligence

Agents
Autonomous ***chatbots***
with access to tools and APIs

Agentic Crews
A team of self-organizing ***Agents***

Simplified Architecture



Our Playbook

Education

- Week of AI
- Threat Modeling
- Prompt Engineering
- Use Cases

Experimentation

- CoPilot
- ChatGPTs
- Hackathons

Platform

- Application Types
- Threat Models
- Model Validation
- Guardrails
- Observability

Summary

- Provide a safe and secure sandbox for your teams to experiment
- ***Optimize*** experiences for non-technical users
- Be generous with observability to cater to variety of users
- Facilitate learning from each other (Shareable Prompts, Catalogs, Templates)
- GenAI applications are naturally indeterministic
- Traditional testing strategies do not work
- It is ***critical*** to put boundaries to generic Foundation Models

"You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete."

R. Buckminster Fuller

SOFTRAMS