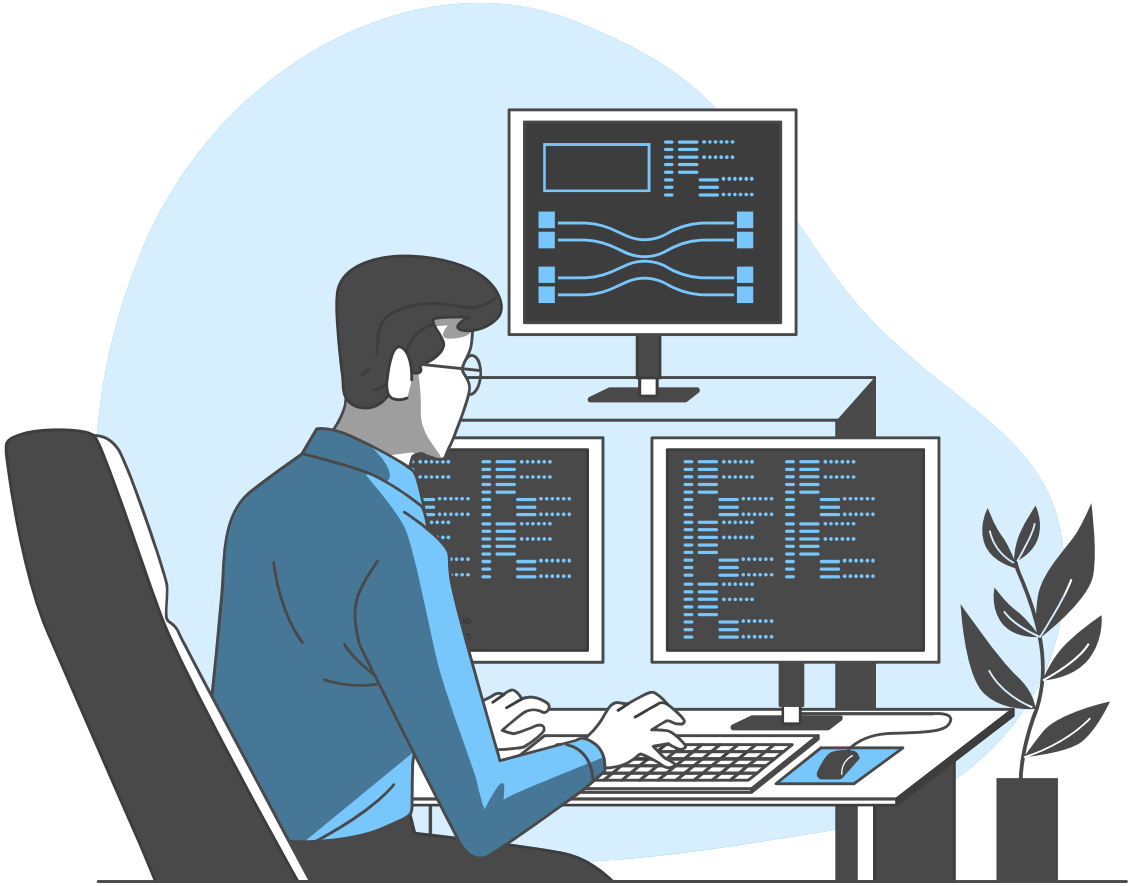
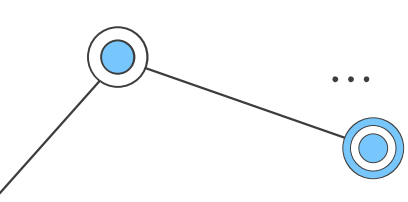
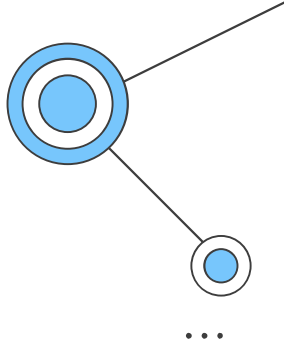


AI Powered Kubernetes

Making our logs easier!

Agenda





01 Introduction

What we will do?

02 Understanding GenAI

Lets learn GenAI

03 Setup

Real stuff

04 QnA

Wanna ask questions?

Who am I?

- **AI/ML Consultant** by day, Chatbot Whisperer by night.
- Currently pursuing an **MCA in AI**—because who needs free time anyway?
- Proud **Rasa Open Source Contributor** and part-time debugger of life's little errors.
- Active **IEEE Young Professional** with a knack for turning jargon into jokes (and then back into code).
- Passionate about building **AI solutions** that are smarter than my fridge—and just as cool.

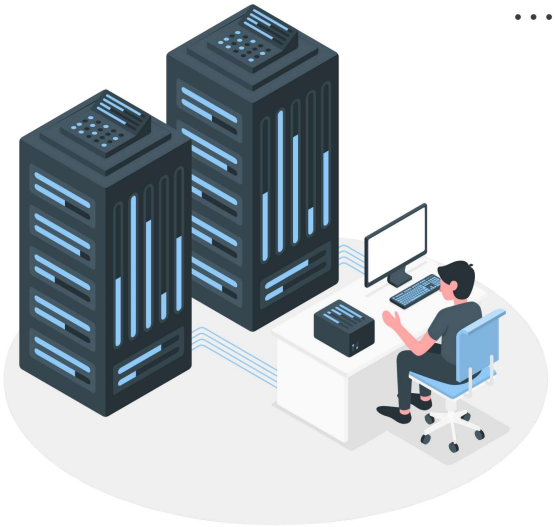
01
Introduction

What we will learn?

Introduction

Unlocking the Future of Kubernetes with Generative AI

Streamlining Kubernetes Management: Harnessing the power of Llama3.1 to optimize Kubernetes automation and scalability.



What You'll Learn:

- Introduction to Generative AI and its role in modern infrastructure.
- How AI can enhance log analysis for faster troubleshooting.
- Hands-on Coding: Using LLMs to understand logs and assist with deployment.

02

What is AI?

AI and Generative AI in Real-World

AI and Generative AI in Real-World

What is AI?

- AI enables machines to mimic human intelligence, including decision-making, learning, and pattern recognition.

Generative AI in Real world

- Co-pilot, that assists devs write code
- Generating weather reports and forecasts.
- Conversational AI for Status updates and observability.

Did You Know?
Generative AI models are also used nowadays for writing final year projects😉

03

RAG

Retrieval Augmented Generation

What the H is RAG?

Definition

- RAG combines retrieval (fetching relevant documents or data) with generation (creating responses based on the retrieved data).

How RAG Works

- AI retrieves the latest information from external or internal sources (e.g., databases, web).
- It uses this data to generate contextually accurate and specific outputs.

04

Let’s Build!

Best creation starts with our dreams

Setting up the environment

- Make sure to have Python 3.10 or above
- Install Anaconda/Miniconda on to your machine
- Install Ollama
- Once Ollama is installed, download Llama 3.2 1B or 3B model as per your machine's capacity
- To Download llama3.2, run *ollama run llama3.2:1b* or *ollama run llama3.2:latest*
- Now, run *conda create -n kube-ai python==3.10* on your command prompt
- Activate your conda environment with *conda activate kubeai*