



KodeKloud

© Copyright KodeKloud

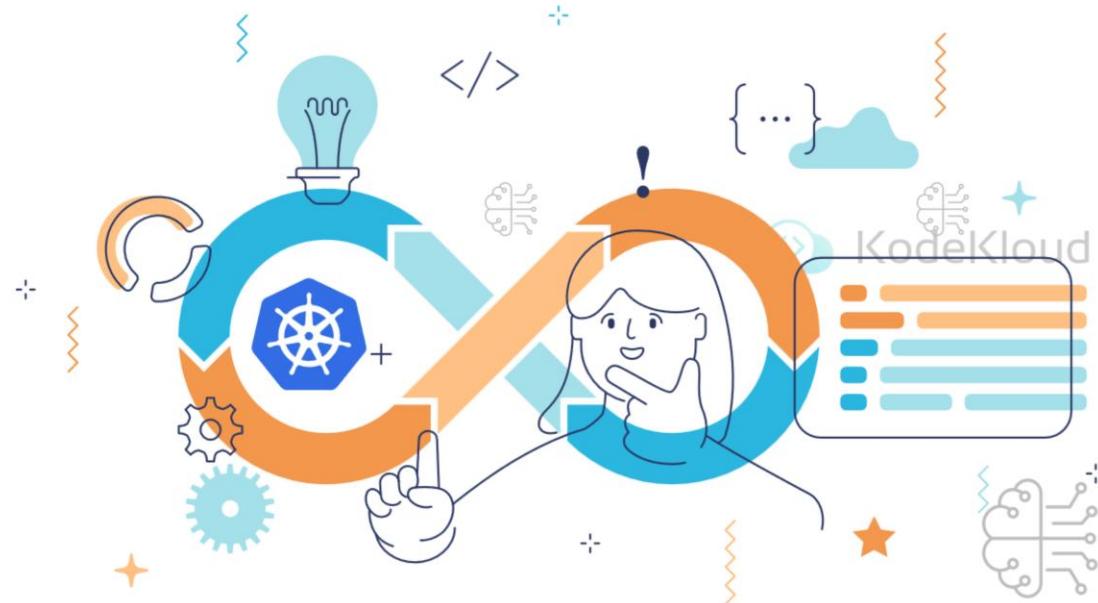
Visit www.kodekloud.com to learn more.

K8sGPT and the rise of AI agents: Navigating the Future of Kubernetes Engineering

© Copyright KodeKloud

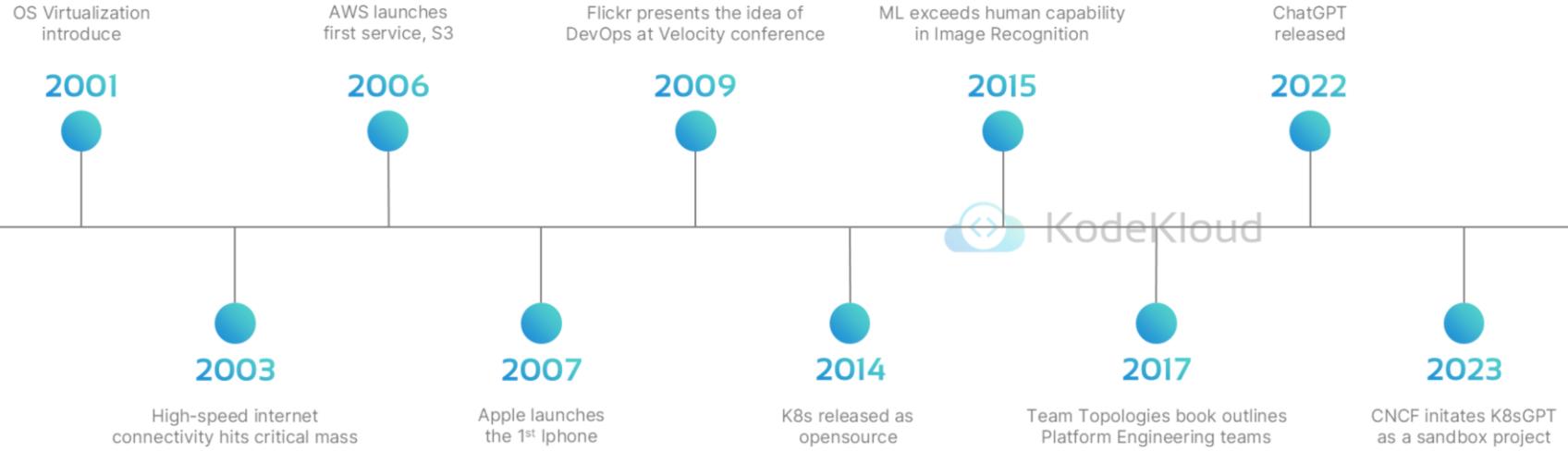
In this course, students will explore how AI is transforming Kubernetes engineering and the DevOps landscape. The speaker (Michael Forrester), an industry veteran, will delve into tools like K8sGPT, which enable autonomous AI agents to optimize and enhance Kubernetes operations, signaling an "agentic future" in DevOps. Participants will gain insights to prepare for the evolving Kubernetes management era driven by AI advancements.

Why are we talking about **Kubernetes** and **AI**?



© Copyright KodeKloud

Why are we talking about this why is this even important what is the intersection between Kubernetes and AI for us to get into this let's talk a little bit about just history that you're all aware of, but I just want to frame this up



© Copyright KodeKloud

Let's look at this timeline:

1970s-1990s: Internet introduced, takes hold in the 90s with dial-up

Early 2000s: Move away from traditional mainframe and server models to commodity server model

2001: VMware introduces ESX desktop for operating system virtualization

2003: High-speed internet connectivity hits critical mass

2006: AWS launches its first cloud service, S3

2007: Apple launches the first iPhone

2008: Android launches (T-Mobile Android T1 in the US)

2009: Flickr presents the idea of DevOps at O'Reilly Velocity Conference

2014: Kubernetes releases as open source

2015: Machine learning matches or exceeds human capabilities in image recognition

2017: "Team Topologies" introduces platform engineering for teams

2022: ChatGPT is released (November)

2023: CNCF ratifies K8sGPT as a Sandbox project (December)

Why are we talking about Kubernetes and AI?

2026

- 30% use digital charisma filters.

2027

- AI productivity as **national power**.
- GenAI cuts legacy app by **70%**.

2028

- **Smart robots** outnumber frontline workers.
- Knowledge worker unionization up **1000%**.

- **GPT-5** interprets most code types.
- **Personalized AI agent's** standard in tech
- Humans oversee AI task management.



© Copyright KodeKloud

By 2026, 30% of workers will leverage digital charisma filters to achieve previously unattainable advances in their career.

By 2027, the productivity value of AI will be recognized as a primary economic indicator of national power.

By 2027, GenAI tools will be used to explain legacy business applications and create appropriate replacements, reducing modernization costs by 70%.

By 2028, there will be more smart robots than frontline workers in manufacturing, retail and logistics due to labor shortages.

By 2028, the rate of unionization among knowledge workers will increase by 1,000%, motivated by the adoption of GenAI.

Inspiration

1.

https://www.gartner.com/doc/reprints?__hstc=254338199.dfb32abfb3189962387122c6e2bbd286.1704311186728.1704311186728.1704311186728.1&__hssc=254338199.8.1704311186728&__hsfp=1654019965&id=1-2FZ2U5YZ&ct=231218&st=sb&submissionGuid=4c21bbc0-5ce2-42d7-a424-851736fc7478

Some additional ones

By 2026, GPT-5 will be out allowing for natural language processing and interpretation of most code generation workloads (YAML, JSON, C++, Helm Charts, etc.)

By 2027, personalized and work-based AI agents will be the defector norm in most technology interaction including work tasks, prioritization, with human beings providing acceptably gates as well as creative oversight.

By 2028, AI will be used by most workers and will have replaced a large percentage of digital implementation jobs.

Our prediction: AI will be used by most workers and will have replaced a large percentage of digital implementation jobs.

This is why we're talking about Kubernetes and AI. Kubernetes was the last layer of innovation that completely disrupted the industry. AI is headed in the same direction. How we innovate on this intersection between Kubernetes, AI, DevOps, and engineering is going to drive a lot of innovation.

Traditional vs AI-assisted Kubernetes Management



© Copyright KodeKloud

The course explores how AI atop Kubernetes is driving a fundamental shift in DevOps, enabling autonomous agents to optimize operations.

Traditional Kubernetes Management



AI-Assisted Kubernetes Management



© Copyright KodeKloud

One of the things to consider is that we're going to move quickly away from our traditional command line to change relationship with Kubernetes to an AI to change relationship there might still be some command line involved but we're going to see an evolution over the next 5 to 10 years where natural language processing is going to be the normal interface for us to change and engineer objects and it runs deeper than that

Key Benefits

01



Simplified diagnostics
and troubleshooting

02



Automated task
execution and
optimization

03



Enhanced decision-
making through
predictive analytics

04



 KodeKloud

Improved security
through intelligent
threat detection

05



Increased efficiency
and speed in software
development and
delivery

© Copyright KodeKloud

Let's start by looking at the key benefits of AI-powered tools in this space: reminder that this is actually just for 2024 we actually don't even know what the improvements are going to be we know that these are the improvements that we're seeing and predicting against now so we don't want to hype this up or overblow it but did anyone actually think in 2006 that cloud was going to be as popular as it was the same with Kubernetes same with platform engineering we didn't know. So, let's look at some benefits.

1. Simplified diagnostics and troubleshooting: AI can quickly analyze vast amounts of log data and system metrics to identify issues that might take humans hours or even days to pinpoint. For example, tools like K8sGPT can scan your entire Kubernetes cluster and provide plain-English explanations of problems it finds.
2. Automated task execution and optimization: AI agents can handle routine tasks like scaling resources, updating configurations, and even performing basic debugging. This frees up human operators to focus on more complex, strategic work.
3. Enhanced decision-making through predictive analytics: By analyzing historical data and current trends, AI can predict potential issues before they occur. This proactive approach can significantly reduce downtime and improve overall system reliability.
4. Improved security through intelligent threat detection: AI-powered security tools can detect anomalies and potential security threats much faster than traditional methods, enhancing the overall security posture of Kubernetes environments.
5. Increased efficiency and speed in software development and delivery: AI tools can assist in various stages of the development pipeline, from code generation and review to testing and deployment, accelerating the entire process.

AI-powered tools

Simplified diagnostics and troubleshooting



splunk>



Automated task execution and optimization



turbonomic

Enhanced decision-making through predictive analytics



Improved security through intelligent threat detection



Increased efficiency in software development and delivery



© Copyright KodeKloud

Here are some software companies and projects related to AI-powered tools in the Kubernetes and DevOps space:
Simplified diagnostics and troubleshooting:

Datadog: Offers AI-powered log management and analytics

Splunk: Provides AI-driven IT Operations Analytics (ITOA)

Dynatrace: Uses AI for application performance management

New Relic: Offers AI-powered observability platform

Automated task execution and optimization:

Red Hat OpenShift: Incorporates AI for automated operations

Mesosphere DC/OS: Uses AI for container orchestration and resource optimization

Turbonomic: Provides AI-powered workload automation for hybrid cloud environments

Enhanced decision-making through predictive analytics:

IBM Watson AIOps: Uses AI for IT operations management and predictive analytics

BMC Helix: Offers AI-driven service management and operations

Moogsoft: Provides AI-powered IT operations management

Improved security through intelligent threat detection:

Darktrace: Uses AI for cyber defense

Vectra AI: Offers AI-driven threat detection and response

ExtraHop Reveal(x): Provides AI-powered network detection and response

Increased efficiency in software development and delivery:

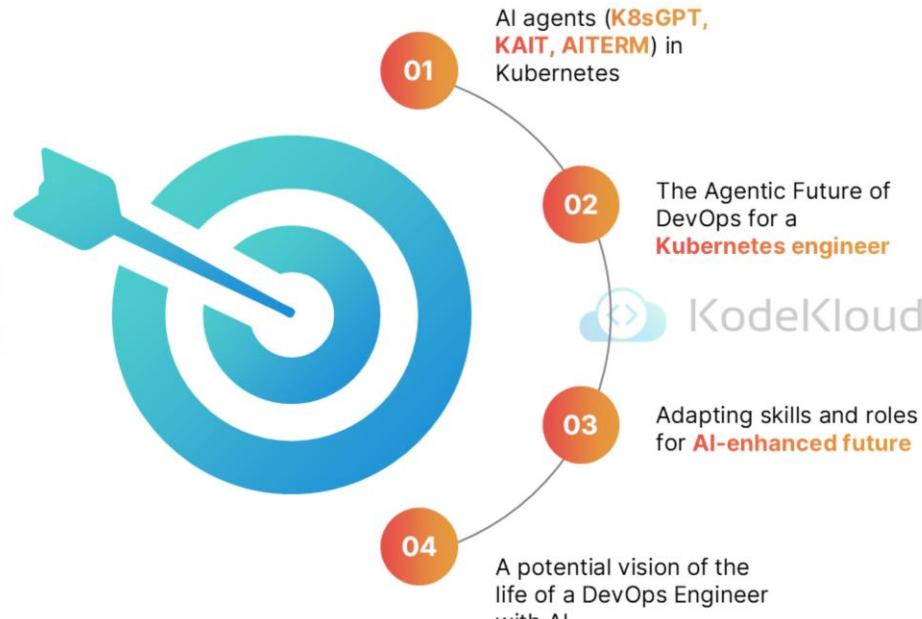
GitHub Copilot: AI-powered code completion tool

DeepCode: AI-based code review and bug detection

Kobiton: AI-powered mobile app testing platform

Functionize: AI-driven test automation platform

What is our goal for this course?



© Copyright KodeKloud

Here are the notes in a concise format:

1. Discuss AI agents and tools approximating AI agents
2. Speculate on an "agentic future" for DevOps and Kubernetes
3. Discuss the foundations needed - shift from implementation to architecture, managing systems without direct intervention
4. Questions and answers on the evolving landscape

Introducing K8sGPT – the
beginning of AI agents





K8sGPT is an open-source tool that scans Kubernetes clusters

© Copyright KodeKloud

Overview:

K8sGPT is an open-source tool that scans Kubernetes clusters, diagnoses issues, and provides explanations in simple English. It's built to address the growing complexity of Kubernetes environments and the challenges that come with managing them at scale.

Key Features

01



Cluster Scanning

02



Issue Diagnosis

03



 KodeKloud
Plain English explanation

04



Actionable advice

05



Anonymization

06



Extensibility

© Copyright KodeKloud

Key Features and Capabilities:

1. Cluster Scanning: K8sGPT can quickly scan your entire Kubernetes cluster, identifying potential issues across various resource types.
2. Issue Diagnosis: Using AI-powered analyzers, K8sGPT can identify problems with pods, services, deployments, and other Kubernetes objects.

3. Plain English Explanations: One of K8sGPT's most powerful features is its ability to explain issues in simple, human-readable language. This makes it accessible even to those who aren't Kubernetes experts.
4. Actionable Advice: Beyond just identifying issues, K8sGPT can provide suggestions on how to fix problems, often including specific kubectl commands to execute.
5. Anonymization: K8sGPT includes a feature to anonymize sensitive data before sending it to AI backends, addressing potential privacy concerns.
6. Extensibility: Users can create custom analyzers to tailor K8sGPT to their specific needs and environments.

Supported AI Backends



© Copyright KodeKloud

Supported AI Backends:

K8sGPT is designed to be flexible in terms of the AI models it can use. Currently supported backends include:

- OpenAI (default)
- Azure OpenAI
- LocalAI (for air-gapped environments)
- Cohere

- Amazon Bedrock
- Google Gemini models

This flexibility allows organizations to choose the AI provider that best fits their needs and compliance requirements.

Demo or screenshots of K8sGPT in action

```
ollama@Bakugo:~/demo-aiterm$ k8sgpt analyse --explain
100% |████████████████████████████████████████████████| (7/7, 1 it/s)
AI Provider: openai

0: Deployment k8sgpt/k8sgpt-ollama()
- Error: Deployment k8sgpt/k8sgpt-ollama has 1 replicas but 0 are available
Error: Deployment k8sgpt/k8sgpt-ollama has 1 replica but 0 are available.

Solution:
1. Check pod status: `kubectl get pods -n k8sgpt`
2. Describe pod for issues: `kubectl describe pod <pod-name> -n k8sgpt`
3. Check logs: `kubectl logs <pod-name> -n k8sgpt`
4. Resolve any errors found in the pod description or logs.

1: Service hosting/web2()
- Error: Service has not ready endpoints, pods: [Pod/web2-878478cf8-vc29x Pod/web2-878478cf8-5w2ck], expected 2
Error: The service has no ready endpoints because the pods are not in a ready state.

Solution:
1. Check pod status with `kubectl get pods`.
2. Inspect pod logs using `kubectl logs <pod-name>`.
3. Verify readiness probes in pod specs.
4. Ensure pods are running and healthy.
5. Restart pods if necessary.

2: Service k8sgpt/k8sgpt-ollama()
- Error: Service has no endpoints, expected label app=k8sgpt-ollama
Error: The service has no endpoints because it cannot find any pods with the label 'app=k8sgpt-ollama'.

Solution:
1. Check if any pods have the label 'app=k8sgpt-ollama'.
2. If not, label the appropriate pods with `kubectl label pod <pod-name> app=k8sgpt-ollama`.
3. Ensure the pods are running and ready.

3: ReplicaSet k8sgpt/k8sgpt-ollama-668c898698()
- Error: pods "k8sgpt-ollama-668c898698-" is forbidden: error looking up service account k8sgpt/k8sgpt: serviceaccount "k8sgpt" not found
Error: The pod "k8sgpt-ollama-668c898698-" is forbidden because the service account "k8sgpt" in the "k8sgpt" namespace was not found.

Solution:
1. Verify the service account exists: `kubectl get serviceaccount k8sgpt -n k8sgpt`
2. If not, create it: `kubectl create serviceaccount k8sgpt -n k8sgpt`
```

© Copyright KodeKloud

Demo:

As you can see, K8sGPT not only identified the issue but also provided a clear explanation and steps to troubleshoot.

K8sGPT represents a significant step towards an Agentic future in DevOps. By automating the analysis of complex Kubernetes environments and providing human-readable explanations, it's changing the way we approach cluster

management and troubleshooting.

In our next slide, we'll explore this concept of an Agentic future in more depth and discuss how tools like K8sGPT are paving the way for a new era in DevOps.

The Agentic Future of DevOps



© Copyright KodeKloud

As we've seen with K8sGPT, Kait, and AIterm, , AI is already making significant inroads in DevOps and Kubernetes management. But what we're witnessing now is just the beginning. We're moving towards what we call an "Agentic future" in DevOps, and that's what we'll explore in this section.



An **Agentic future in DevOps** involves AI agents autonomously

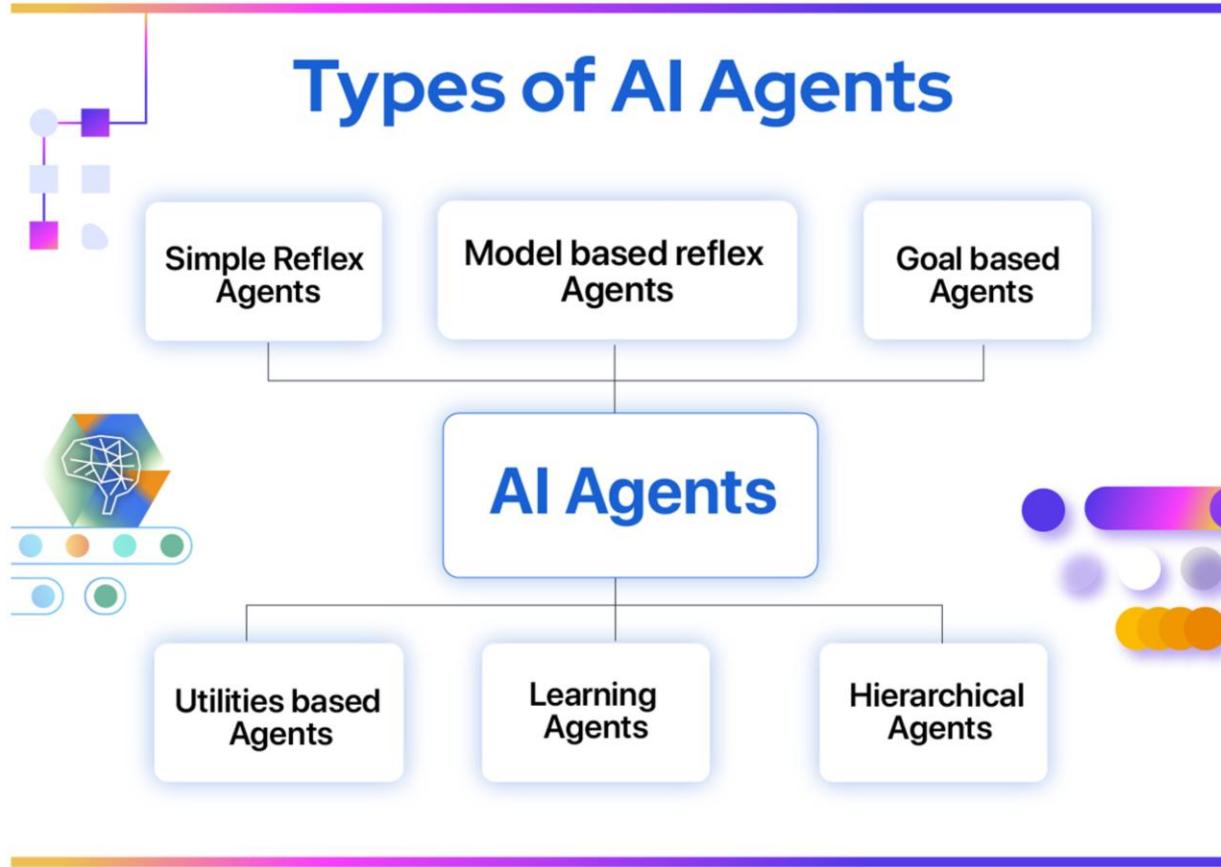
- managing tasks,
 - making decisions, and
 - learning from experiences in Kubernetes and DevOps processes,
- acting as active participants rather than mere tools.



© Copyright KodeKloud

Definition of Agentic AI in DevOps:

An Agentic future in DevOps refers to a scenario where AI agents can autonomously perform complex tasks, make decisions, and even learn from their experiences in managing Kubernetes clusters and other DevOps processes. These AI agents are not just tools; they're active participants in the DevOps workflow, capable of initiating actions and adapting to new situations.



© Copyright KodeKloud

Imagine very quickly that we're going to get to a world where you start a new job and instantly that job provides a job-based AI agent who talks to your personal AI agent, which is attached to you, and gets your preferences and settings for how you like to work. You know, how you like to be talked to, what your language is, accents—all this stuff is for your personal natural language interface.

Let's say that my work-based agent—let's call it Kodi with a K. So, Kodi is my personal Code Cloud agent for the work that I do with Code Cloud. Kodi works with me on all the projects that I do, whether it's a curriculum-based project for training or

an engineering-based project where I have to put together a proof of concept or some kind of scale project. Kodi's job is to keep me in line with Code Cloud's other agents, which manage security, operations, and compliance.

So, whenever I ask Kodi a question like, "Hey Kodi, can you go out and launch a triple-tier application based on traditional triple-tier architecture? I want you to run it on AWS and pull it from Project Unicorn."

What's happened here is that now Kodi is going to talk to all of the agents that Code Cloud runs around architecture, security standards, and compliance standards, in addition to access and authorization, to make sure that it can pull everything it needs from the repository, build from the repository, run the CI/CD pipeline, build the infrastructure based on the best of triple-tier architecture, and then deploy the app into that environment.

It doesn't matter whether I said I want virtual machines, Kubernetes and containers, or serverless. Honestly, I'm probably not even going to specify. Kodi is going to take a look at the code, the way it's structured, see if it's a monolith versus a microservice, and see what kind of resources it needs because it's actually going to pull it and test it. Then, it's going to choose the best architecture because it has plenty of agents to talk to that will give it advice about how these things should run.

Notice that this all started from one comment by me: "Hey, I wanna run a triple-tier web/app/db architecture."

Key Features

01



Autonomous
problem-
solving

02



Continuous learning
and improvement

03



Predictive and
Proactive

04



KodeKloud
Intelligent
Collaboration

05



Cross system
optimization

© Copyright KodeKloud

Key Features of an Agentic DevOps Future:

1. Autonomous Problem-Solving: AI agents will be able to identify issues, diagnose root causes, and implement solutions without constant human oversight. For example, an Agentic AI might detect a performance bottleneck, analyze its cause, and automatically scale resources or optimize configurations to resolve it.

2. Continuous Learning and Improvement: These AI systems will learn from every interaction and outcome, continuously improving their performance and adapting to the specific needs of your environment. They'll build up a knowledge base that combines general best practices with the unique characteristics of your infrastructure.
3. Predictive and Proactive Management: Going beyond reactive problem-solving, Agentic AI will predict potential issues before they occur and take preventive actions. This could include predicting resource needs and scaling preemptively, or identifying code patterns likely to lead to future bugs.
4. Intelligent Collaboration: Agentic AI will work alongside human DevOps engineers, understanding context, interpreting natural language instructions, and even engaging in dialogue to clarify requirements or explain its actions.
5. Cross-System Optimization: These AI agents will have a holistic view of the entire DevOps pipeline, optimizing across different systems and stages of the development lifecycle for overall efficiency.

Potential impact on DevOps roles and processes

01



Shift in Human
roles

02



Accelerated
Innovation

03



Enhanced reliability
and performance

04



KodeKloud

Democratization of
DevOps

© Copyright KodeKloud

Potential Impact on DevOps Roles and Processes:

1. Shift in Human Roles: DevOps engineers will evolve from hands-on operators to strategic overseers and AI trainers. They'll focus on setting high-level goals, defining policies, and handling complex edge cases that AI can't manage.
2. Accelerated Innovation: With routine tasks automated, teams can focus more on innovation and high-value activities.

This could lead to faster feature development and more rapid evolution of applications.

3. Enhanced Reliability and Performance: Continuous, AI-driven optimization could lead to significantly improved system reliability and performance, with fewer human-induced errors.

4. Democratization of DevOps: As AI makes complex operations more accessible, we might see a broadening of who can effectively manage and deploy complex systems.

Demo or screenshots of KAIT in action

```
ollama@Bakugo:~/demo-kait$ kait debug "A service in the hosting Namespace is not responding to requests. Determine which Service is not working and resolving the underlying issue so the Service begins responding to requests."  
Let's start by listing all the services in the 'hosting' namespace to identify which one might be having issues.
```

```
```sh  
kubectl get services -n hosting
```
```

| NAME | TYPE | CLUSTER-IP | EXTERNAL-IP | PORT(S) | AGE |
|------|----------|---------------|-------------|--------------|-------|
| web1 | NodePort | 10.103.16.243 | <none> | 80:31329/TCP | 2m56s |
| web2 | NodePort | 10.110.61.229 | <none> | 80:32459/TCP | 2m56s |

Next, let's check the endpoints for these services to see if they are correctly pointing to the pods.

```
```sh  
kubectl get endpoints -n hosting
```
```

| NAME | ENDPOINTS | AGE |
|------|---------------------------|-------|
| web1 | 10.1.0.26:80,10.1.0.27:80 | 2m57s |
| web2 | | 2m57s |

It looks like the 'web2' service does not have any endpoints, which means it is not pointing to any pods. Let's check the pods in the 'hosting' namespace to see if there are any pods that should be associated with the 'web2' service.

```
```sh  
kubectl get pods -n hosting
```
```

| NAME | READY | STATUS | RESTARTS | AGE |
|-----------------------|-------|---------|----------|-------|
| web1-8586dbfcf9-4x8cn | 1/1 | Running | 0 | 2m58s |
| web1-8586dbfcf9-jh7jr | 1/1 | Running | 0 | 2m58s |
| web2-878478cfcd-27vzs | 0/1 | Running | 0 | 2m58s |
| web2-878478cfcd-z25rm | 0/1 | Running | 0 | 2m58s |

The 'web2' pods are not in a ready state, which is likely why the 'web2' service has no endpoints. Let's describe one of the 'web2' pods to get more information about why it is not ready.

© Copyright KodeKloud

OK so here you're going to do a live demo of something called KAIT. This is a tool that allows you to debug a Kubernetes cluster just by saying hey this is the problem that I'm seeing whether your application isn't being exposed whether you are seeing you know some kind of image problem. This tool basically allows you to identify quickly what's going on and it will actually show you the commands it'll start by listing the services and everything else there is a mode for this where it is read only set to false. And in that mode that allows you to actually correct issues. Let's take a deeper look at this live

Emerging Trends

01



AI-Driven
Observability

02



Autonomous
Kubernetes Interface

03



Natural Language
Interface

04



Predictive
Coding

KodeKloud

© Copyright KodeKloud

Emerging Trends:

1. **AI-Driven Observability:** Moving beyond monitoring to provide deeper insights into system behavior and performance.
2. **Autonomous Kubernetes Management:** AI systems that can manage complex Kubernetes environments with minimal human intervention.

3. Natural Language Interfaces: DevOps tools that can understand and execute commands given in natural language.
4. Predictive Coding: AI systems that can generate entire applications based on high-level specifications.

The Future of DevOps Engineers in an AI-Driven World



© Copyright KodeKloud

As we've explored the rise of AI in Kubernetes and DevOps, it's clear that the role of DevOps engineers is set to undergo significant changes. In this section, we'll discuss how these roles are likely to evolve and what skills will be crucial for success in this new landscape.

Evolving Role of DevOps Engineers



DevOps engineers will shift from directly executing tasks to orchestrating AI systems that perform these tasks.

They'll focus more on designing workflows, setting policies, and defining high-level strategies.

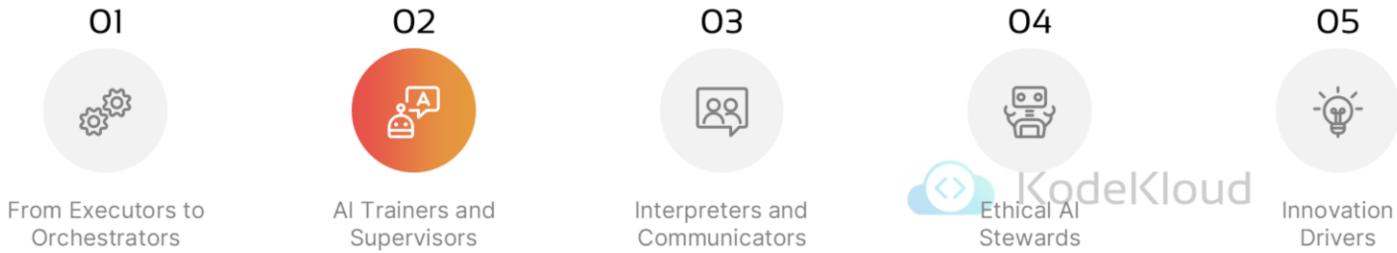
© Copyright KodeKloud

Evolving Role of DevOps Engineers:

1. From Executors to Orchestrators:

- DevOps engineers will shift from directly executing tasks to orchestrating AI systems that perform these tasks.
- They'll focus more on designing workflows, setting policies, and defining high-level strategies.

Evolving Role of DevOps Engineers



© Copyright KodeKloud

Evolving Role of DevOps Engineers:

2. AI Trainers and Supervisors:

- A key responsibility will be training AI systems on organization-specific practices and requirements.
- Engineers will need to supervise AI decisions, especially in critical or novel situations.

Evolving Role of DevOps Engineers



As AI systems become more complex, DevOps engineers will play a crucial role in interpreting AI insights for other stakeholders.

They'll need to effectively communicate between AI systems and human team members.

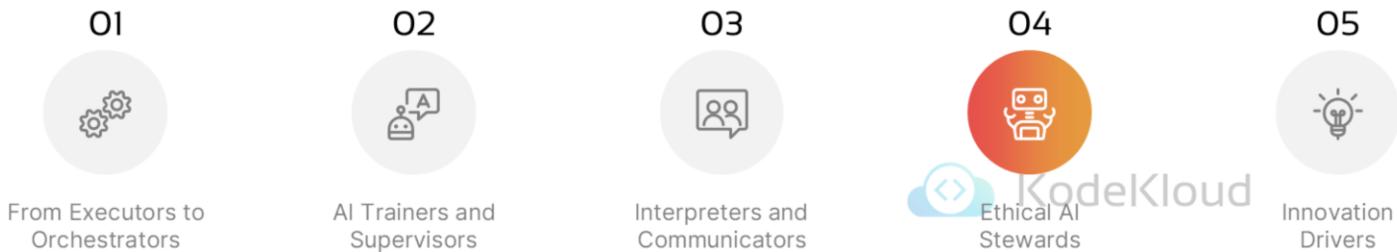
© Copyright KodeKloud

Evolving Role of DevOps Engineers:

3. Interpreters and Communicators:

- As AI systems become more complex, DevOps engineers will play a crucial role in interpreting AI insights for other stakeholders.
- They'll need to effectively communicate between AI systems and human team members.

Evolving Role of DevOps Engineers



DevOps engineers will be responsible for ensuring AI systems are used ethically and in compliance with regulations.

They'll need to monitor for bias and fairness in AI-driven processes.

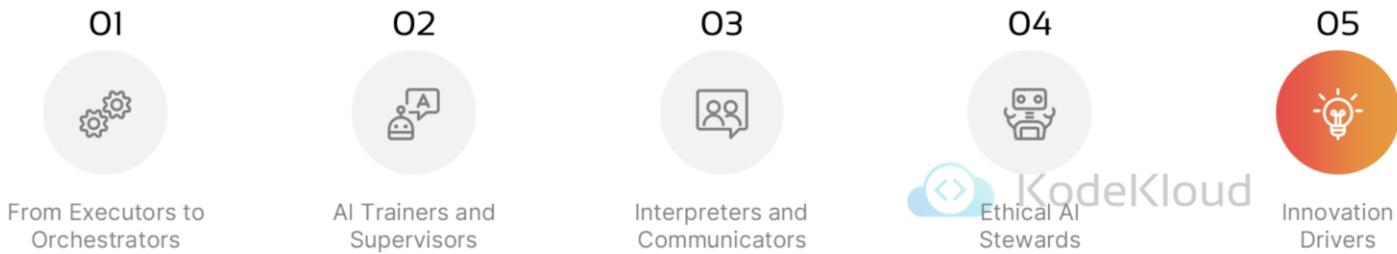
© Copyright KodeKloud

Evolving Role of DevOps Engineers:

4. Ethical AI Stewards:

- DevOps engineers will be responsible for ensuring AI systems are used ethically and in compliance with regulations.
- They'll need to monitor for bias and fairness in AI-driven processes.

Evolving Role of DevOps Engineers



With routine tasks automated, DevOps engineers will have more capacity to focus on innovation and strategic improvements.

© Copyright KodeKloud

Evolving Role of DevOps Engineers:

5. Innovation Drivers:

- With routine tasks automated, DevOps engineers will have more capacity to focus on innovation and strategic improvements.

New Skills and Competencies Required

01



AI and Machine
Learning
Fundamentals

02



Data Engineering

03



Advanced
Programming

04



AI Ethics and
Governance

05



Systems
Thinking

06



Soft Skills

Understanding of ML algorithms, neural networks,
and natural language processing.

Ability to work with and fine-tune AI models.

© Copyright KodeKloud

New Skills and Competencies Required:

1. AI and Machine Learning Fundamentals:

- Understanding of ML algorithms, neural networks, and natural language processing.
- Ability to work with and fine-tune AI models.

New Skills and Competencies Required

01



AI and Machine
Learning
Fundamentals

02



Data Science and
Analytics

03



Advanced
Programming

04



AI Ethics and
Governance

05



Systems
Thinking

06



Soft Skills

Skills in data preparation, feature engineering, and statistical analysis.

Proficiency in data visualization to communicate AI insights effectively.

© Copyright KodeKloud

New Skills and Competencies Required:

2. Data Science and Analytics:

- Skills in data preparation, feature engineering, and statistical analysis.
- Proficiency in data visualization to communicate AI insights effectively.

New Skills and Competencies Required

01



AI and Machine
Learning
Fundamentals

02



Data Science and
Analytics

03



Advanced
Programming

04



AI Ethics and
Governance

05



Systems
Thinking

06



Soft Skills

Beyond traditional scripting, knowledge of languages commonly used in AI/ML (e.g., Python, R).

Understanding of AI frameworks and libraries (e.g., TensorFlow, PyTorch).

© Copyright KodeKloud

New Skills and Competencies Required:

3. Advanced Programming:

- Beyond traditional scripting, knowledge of languages commonly used in AI/ML (e.g., Python, R).
- Understanding of AI frameworks and libraries (e.g., TensorFlow, PyTorch).

New Skills and Competencies Required

01



AI and Machine
Learning
Fundamentals

02



Data Science and
Analytics

03



Advanced
Programming

04



AI Ethics and
Governance

05



Systems
Thinking

06



Soft Skills

Knowledge of ethical AI principles and their practical application.

Understanding of AI-related regulations and compliance requirements.

© Copyright KodeKloud

New Skills and Competencies Required:

4. AI Ethics and Governance:

- Knowledge of ethical AI principles and their practical application.
- Understanding of AI-related regulations and compliance requirements.

New Skills and Competencies Required

01



AI and Machine
Learning
Fundamentals

02



Data Science and
Analytics

03



Advanced
Programming

04



AI Ethics and
Governance

05



KodeKloud
Systems
Thinking

06



Soft Skills

Ability to understand and optimize complex, AI-driven systems.

Skills in designing resilient, self-healing architectures.

© Copyright KodeKloud

New Skills and Competencies Required:

5. Systems Thinking:

- Ability to understand and optimize complex, AI-driven systems.
- Skills in designing resilient, self-healing architectures.

New Skills and Competencies Required

01



AI and Machine
Learning
Fundamentals

02



Data Science and
Analytics

03



Advanced
Programming

04



AI Ethics and
Governance

05



Systems
Thinking

06



Soft Skills

Enhanced communication skills to explain AI concepts to non-technical stakeholders.

Adaptability and continuous learning mindset.

Creativity for problem-solving in AI-human hybrid environments.

© Copyright KodeKloud

New Skills and Competencies Required:

6. Soft Skills:

- Enhanced communication skills to explain AI concepts to non-technical stakeholders.
- Adaptability and continuous learning mindset.
- Creativity for problem-solving in AI-human hybrid environments.

Emerging Job Roles in AI-Driven DevOps



AI Ops Specialist



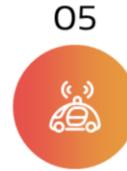
DevOps AI Trainer



AI Ethics Officer



DevOps Data Scientist



Autonomous Systems Architect



© Copyright KodeKloud

Emerging Job Roles in AI-Driven DevOps:

1. AI Ops Specialist:

- Focuses on integrating and managing AI systems within DevOps workflows.

2. DevOps AI Trainer:

- Specializes in training and fine-tuning AI models for specific DevOps use cases.

3. AI Ethics Officer:

- Ensures ethical use of AI in DevOps processes and compliance with regulations.

4. DevOps Data Scientist:

- Bridges the gap between data science and DevOps, focusing on ML model deployment and management.

5. Autonomous Systems Architect:

- Designs self-managing, AI-driven infrastructure and application architectures.

Overview of AI Applications in DevOps

01



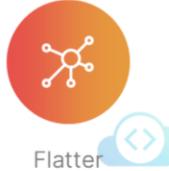
Cross-functional
AI Teams

02



AI Center of
Excellence

03



Flatter
Hierarchies

04



KodeKloud

Global
Collaboration

© Copyright KodeKloud

Impact on Team Structures and Collaboration:

1. Cross-functional AI Teams:

- DevOps engineers will increasingly work in teams that include data scientists and AI specialists.

2. AI Center of Excellence:

- Many organizations may establish central teams to guide AI adoption across DevOps and other functions.

3. Flatter Hierarchies:

- As AI takes over routine tasks, team structures may become flatter, with a focus on expertise rather than traditional hierarchies.

4. Global Collaboration:

- AI tools may enable more effective collaboration across geographically dispersed teams.

Demo or screenshots of AITERM in action

```
ollama@Bakugo:~/demo-k8sgpt$ aiterm troubleshoot my current Nginx deployment and correct whatever is wrong with it to get it back to running
Commands needed to run:
• kubectl get deployment nginx
• kubectl describe deployment nginx
• kubectl get pods --selector=app=nginx
• kubectl logs deployment/nginx
• kubectl rollout status deployment/nginx
• kubectl rollout restart deployment/nginx

? Do you want to run all the commands? Yes
kubectl get deployment nginx
NAME      READY   UP-TO-DATE   AVAILABLE   AGE
nginx    1/1     1           1           3h8m
kubectl describe deployment nginx
Name:                  nginx
Namespace:             default
CreationTimestamp:     Tue, 02 Jul 2024 16:41:13 -0400
Labels:                app=nginx
Annotations:           deployment.kubernetes.io/revision: 2
Selector:              app=nginx
Replicas:              1 desired | 1 updated | 1 total | 1 available | 0 unavailable
StrategyType:          RollingUpdate
MinReadySeconds:       0
RollingUpdateStrategy: 25% max unavailable, 25% max surge
Pod Template:
  Labels:        app=nginx
  Annotations:  kubectl.kubernetes.io/restartedAt: 2024-07-02T19:48:48-04:00
  Containers:
    nginx:
      Image:      nginx
      Port:       <none>
      Host Port: <none>
      Environment: <none>
      Mounts:    <none>
      Volumes:   <none>
  Conditions:
    Type     Status  Reason
    ----   -----  -----
    Available  True    MinimumReplicasAvailable
    Progressing  True    NewReplicaSetAvailable
OldReplicaSets: nginx-7854ff8877 (0/0 replicas created)
NewReplicaSet:  nginx-bd4d68d68 (1/1 replicas created)
Events:
  Type     Reason     Age   From            Message
  ----   -----     --   --              --
```

© Copyright KodeKloud

- Introduction to AITERM tool
- AITERM integrates AI capabilities into the command line
- Not limited to Kubernetes, can be used for various tasks (Jenkins, Terraform, Raspberry Pi troubleshooting)
- Live demo of AITERM troubleshooting an Nginx deployment in a Kubernetes cluster
- AITERM analyzes the deployment, runs necessary commands, and corrects the issue (wrong image tag/version)
- AITERM can be used to generate code (e.g., Jenkins files, Terraform) without starting from scratch

- AITERM currently uses OpenAI or Llama APIs for the AI capabilities
- This is just a simple example of an "agented world" - more advanced agent-based systems are coming in the future
- The goal is to make these AI-powered capabilities accessible to beginners, though the full agented environment setup can be more complex currently

Preparing for an AI-Driven Kubernetes Future



© Copyright KodeKloud

As we conclude our exploration of K8sGPT and the rise of AI agents in Kubernetes engineering, let's recap the key points and discuss how both individuals and organizations can prepare for this AI-driven future.

Key Features

01



AI is transforming Kubernetes and DevOps, enhancing efficiency and decision-making.

02



Tools like K8sGPT enable AI to autonomously perform complex tasks.

03



DevOps engineers' roles are shifting to strategic orchestration of AI-driven systems.

04



AI introduces challenges like data privacy, skill gaps, and ethical issues.

05



Future DevOps will feature a symbiotic relationship between humans and AI.

© Copyright KodeKloud

Recap of Key Points:

1. AI is transforming Kubernetes management and DevOps practices, offering benefits such as increased efficiency, improved problem-solving, and enhanced decision-making.
2. Tools like K8sGPT are paving the way for an Agentic future in DevOps, where AI can autonomously perform complex

tasks and make decisions.

3. The role of DevOps engineers is evolving from hands-on executors to strategic orchestrators of AI-driven systems.
4. While AI offers significant benefits, it also brings challenges related to data privacy, skill gaps, and ethical considerations.
5. The future of DevOps will likely involve a symbiotic relationship between human engineers and AI systems, each leveraging their unique strengths.

Skills and Knowledge needed for Future Kubernetes Engineers



© Copyright KodeKloud

Skills and Knowledge Needed for Future Kubernetes Engineers:

1. AI and Machine Learning Fundamentals:

- Understanding of ML algorithms, neural networks, and their applications in Kubernetes.
- Ability to work with and fine-tune AI models for specific use cases.

2. Advanced Kubernetes Knowledge:

- Deep understanding of Kubernetes internals and advanced concepts.
- Expertise in Kubernetes operators and custom resource definitions (CRDs).

3. Programming and Scripting:

- Proficiency in languages commonly used in AI/ML and Kubernetes automation (e.g., Python, Go).
- Understanding of AI frameworks and libraries relevant to Kubernetes management.

4. Data Science and Analytics:

- Skills in data preparation and analysis to support AI-driven decision making.
- Ability to interpret and act on AI-generated insights.

5. Systems Architecture:

- Capability to design scalable, resilient, and AI-augmented Kubernetes architectures.
- Understanding of how to integrate AI components into existing Kubernetes environments.

6. Ethics and Governance:

- Knowledge of ethical AI principles and their application in Kubernetes management.
- Understanding of relevant regulations and compliance requirements.

7. Soft Skills:

- Enhanced communication skills to explain complex AI-Kubernetes interactions to various stakeholders.
- Adaptability and a strong commitment to continuous learning.
- Critical thinking and problem-solving in AI-human collaborative environments.

Resources for Further Learning and Development

Online Courses

- Coursera's [AI for Everyone by Andrew Ng](#)
- edX's [Artificial Intelligence for Kubernetes](#)

Books

- [AI-Assisted Python Programming](#) by Leo Porter and Daniel Zingaro
- [Artificial Intelligence: A Modern Approach](#) by Stuart Russell and Peter Norvig

Communities & Forums

- [Kubernetes Slack community](#)
- [AI for DevOps forums](#) on Stack Overflow

Conferences and Webinars

- [KubeCon + CloudNativeCon](#)



Hands-on Practice

- Set up a [K8sGPT](#) environment in a test cluster
- Participate in [AI-focused Kubernetes](#) projects on GitHub

© Copyright KodeKloud

Resources for Further Learning and Development:

1. Online Courses:

- Coursera's "AI for Everyone" by Andrew Ng
- edX's "Artificial Intelligence for Kubernetes"

2. Books:

- "Kubernetes: Up and Running" by Kelsey Hightower et al.
- "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig

3. Communities and Forums:

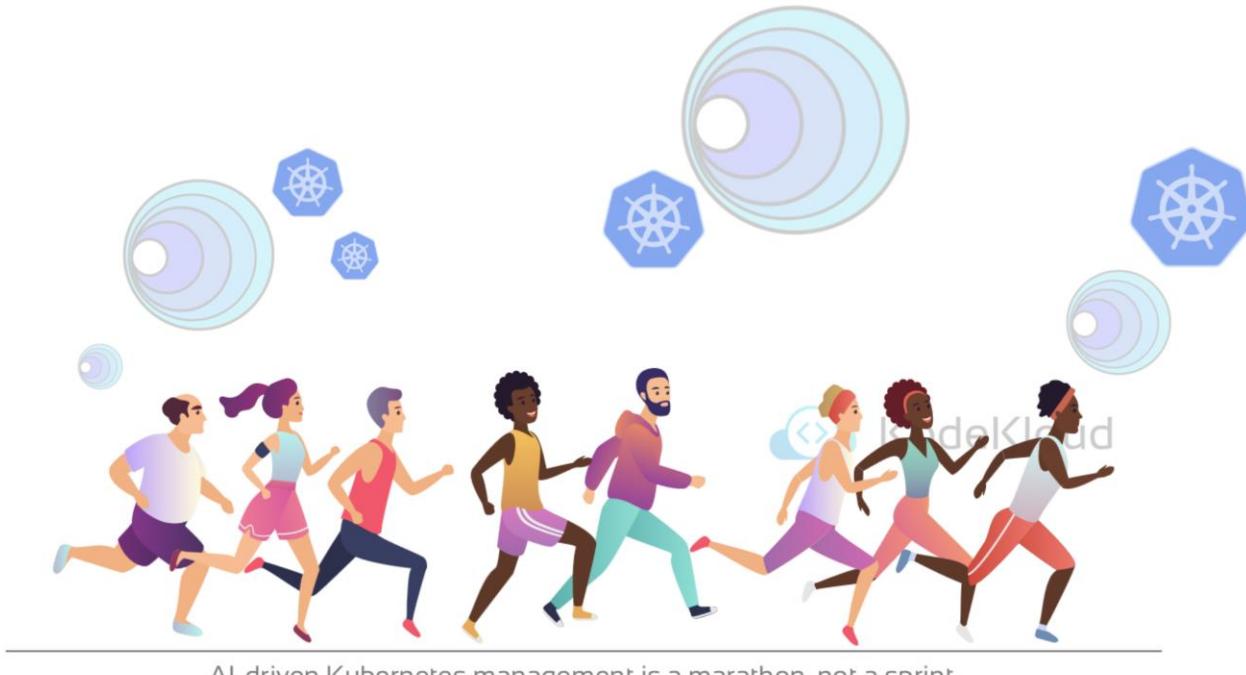
- Kubernetes Slack community
- AI for DevOps forums on Stack Overflow

4. Conferences and Webinars:

- KubeCon + CloudNativeCon
- AI DevOps Summit

5. Hands-on Practice:

- Set up a K8sGPT environment in a test cluster
- Participate in AI-focused Kubernetes projects on GitHub



AI-driven Kubernetes management is a marathon, not a sprint.

© Copyright KodeKloud

Remember, the journey to AI-driven Kubernetes management is a marathon, not a sprint. It requires continuous learning, adaptation, and a willingness to embrace change. By starting now and taking deliberate steps, you can position yourself and your organization at the forefront of this exciting transformation in Kubernetes engineering.

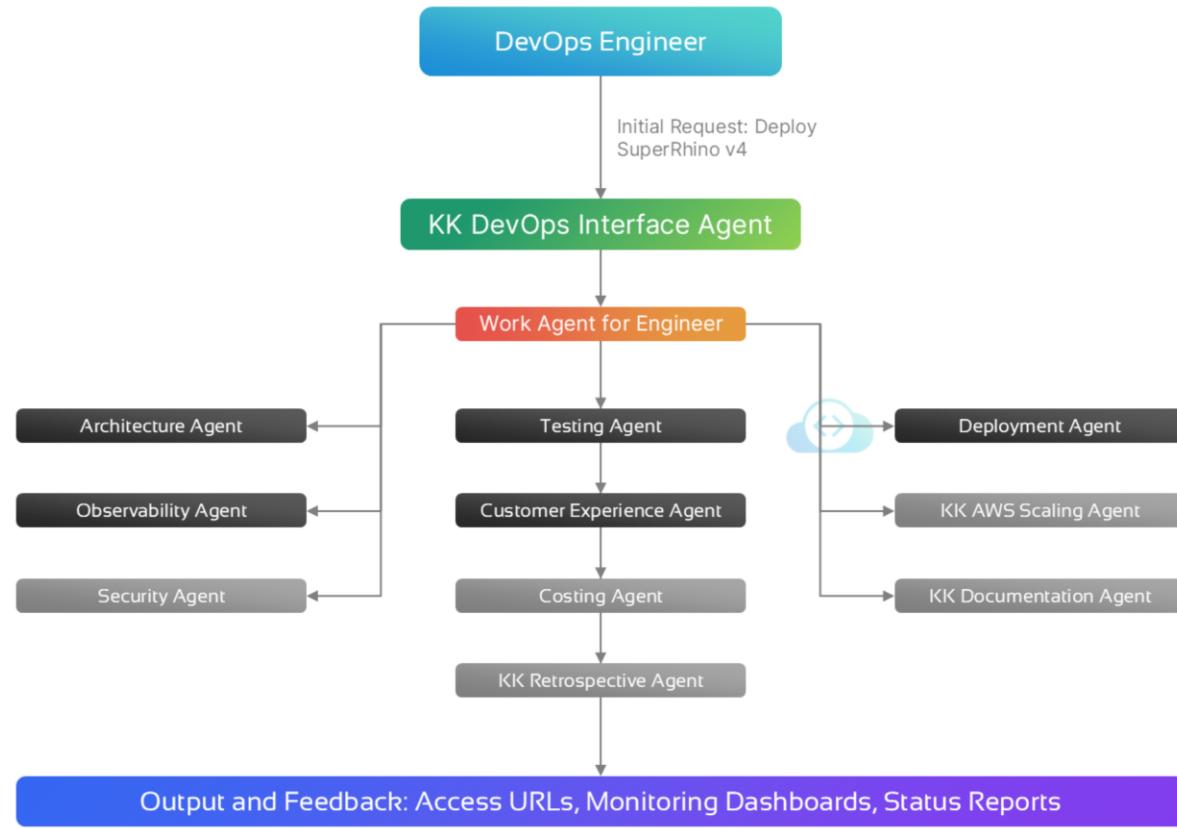
Thank you for your attention throughout this course.

A Multi-Agentic Future for **DevOps Engineers**

How would a request flow from user to assistant to the many agents
in the background?



KodeKloud



© Copyright KodeKloud

● Specialized Agents ● Auxiliary Agents

KK DevOps Interface Agent:

Initial Request

DevOps Engineer: "I need to deploy SuperRhino Application v4 using a three-tier architecture on AWS EKS. This is for production, so I want full monitoring and A/B testing for the final deploy."

Work Agent for Engineer

Acknowledges the request and begins coordinating with specialized agents.

Acts as the primary interface between the DevOps Engineer and other AI agents.

Architecture and Standards Verification

Architecture Agent provides the optimal template for a 3-tier app.

Architecture Agent analyzes the code repository:

- Checks main branch, deployability, test status, and third-party library issues.

- Verifies with Artifact Repository to avoid unnecessary rebuilds.

- Reports back to Work Agent for Engineeron code readiness and artifact availability.

Pre-Production Testing

Testing Agent:

- Runs end-to-end QA tests on the artifact.

- Conducts performance testing in a production-like environment.

- Reports results to KK DevOps Interface Agent, ensuring compliance with KK Standards.

Production Deployment Preparation

Deployment Agent:

- Prepares or identifies an existing EKS cluster.

- Sources container images from the Artifact Repository.

- Coordinates with AWS RDS Agent for database setup (as per KK standards for small SQL databases) (not shown)

Observability Setup

Observability Agent:

- Deploys monitoring stack (Prometheus, Grafana, etc.).

- Ensures alignment with Architecture Agent.

Deployment and A/B Testing

Customer Experience Agent:

- Initiates deployment with 95% traffic to old version, 5% to new version.

- Monitors user experience metrics.

- Gradually increases traffic to the new version based on positive metrics.

Deployment Confirmation

Work Agent for Engineer:

- Confirms successful deployment.
- Provides access URL to User.
- Sends notification email to relevant stakeholders.

2nd Request from User

DevOps Engineer: "We need to add business continuity and scale to 50,000 concurrent users at 12pm on April 15th."

Business Continuity and Scaling

Work Agent for Engineer consults with:

- Architecture Agent for BCP standards.
- Deployment Agent for auto-scaling policies.

KK AWS Scaling Agent:

- Configures auto-scaling to handle 50K users on April 15th at 12pm.
- Sets up CloudWatch alarms and scaling policies.

Security and Compliance Check

Security Agent:

- Verifies that all deployments meet security standards.
- Ensures compliance with relevant regulations (e.g., GDPR, HIPAA).

Cost Optimization

Costing Agent:

- Analyzes the deployment for potential cost optimizations.
- Suggests improvements like using Spot Instances or adjusting resource allocations.

Documentation and Knowledge Base Update

KK Documentation Agent:

- Updates internal wikis and runbooks with new deployment information.
- Generates a summary report of the entire process.

Continuous Improvement

KK Retrospective Agent:

- Analyzes the deployment process for areas of improvement.

Suggests updates to templates and standards based on this deployment.

Human Oversight and Approval

Work Agent for Engineer presents the deployment plan and progress to the DevOps Engineer.

DevOps Engineer reviews the process, approves critical decisions, and provides any necessary adjustments.

Final Update Compilation and Presentation

Work Agent for Engineer:

Compiles a comprehensive deployment update for SuperRhino Application v4

Includes URLs for accessing the application and monitoring dashboards

Provides a status report on the three-tier architecture deployment on AWS EKS

Summarizes the implementation of full monitoring and A/B testing

Confirms the setup for scaling to 50,000 concurrent users on April 15th at 12pm

Presents the deployment results to the DevOps Engineer, including:

- Application access URL

- Monitoring dashboard URL

- Current status of the deployment

- Summary of implemented features (three-tier architecture, monitoring, A/B testing)

- Confirmation of scaling setup

Offers the DevOps Engineer an option to ask follow-up questions or request further actions

DevOps Engineer Acknowledgment and Feedback

DevOps Engineer reviews the deployment information

Provides feedback or asks additional questions if necessary

Confirms satisfaction with the deployment or requests further modifications



DevOps



Scripting

Do we still need to know....



K8s

© Copyright KodeKloud

AI is not yet advanced enough to handle tasks like implementation, architecture, business alignment, or humanizing solutions independently, making human expertise crucial. Domain experts play a key role in training, filtering, and refining AI to ensure it becomes faster, more effective, and business-relevant. Even for beginners, a strong foundation in current skills is essential to remain adaptable and relevant as technology evolves. Understanding the fundamentals of systems—be it Kubernetes, CI/CD pipelines, or programming—is akin to knowing how a car functions; it helps in evaluating, solving, or delegating tasks effectively. To prepare for the future, professionals should aim to be T-shaped, combining broad

knowledge across various domains with deep expertise in one area. This balance ensures both immediate and long-term success in a tech-driven world.



DevOps



Scripting

It will be more important than ever....



K8s

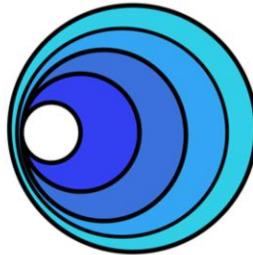


© Copyright KodeKloud

In the coming years, as AI continues to evolve, it will become increasingly important to understand the foundational concepts and architecture patterns of current technologies, especially in the early stages of AI development. While syntax, command-line flags, and implementation details will become less critical, the focus will shift to skills like architecture, providing business value, troubleshooting, and effectively working with AI—becoming caretakers and orchestrators of AI systems. While AI agents will partner with humans to push the boundaries of experimentation and innovation, mastering technologies like Kubernetes, serverless architectures, and understanding large models (LMs) will remain essential.

DevOps professionals will continue to thrive by managing and nurturing these AI agents, ensuring they operate effectively within the broader system.

Conclusion



KodeKloud

© Copyright KodeKloud

The rise of AI agents in Kubernetes and DevOps represents a transformative shift in cloud-native engineering. Tools like K8sGPT are the first steps towards AI systems that not only analyze Kubernetes clusters but also make autonomous decisions to optimize performance, enhance security, and predict potential issues. However, AI is not replacing engineers—it's augmenting them, leaving human creativity, complex problem-solving, and ethical decision-making indispensable.

As implementation increasingly shifts to AI agents, mastering the fundamentals of engineering, architecture, and AI/ML

operations will become essential. DevOps professionals will play a critical role in managing and maintaining these agents across industries like finance, healthcare, and HR. In the next decade, hands-on tasks will diminish, and the focus will pivot towards orchestrating and "care and feeding" of AI agents, as we move toward a fully agent-driven ecosystem.

Key Takeaways

- ✓ AI is transforming **Kubernetes management**, offering improved efficiency, accuracy, and insight.
- ✓ Tools like **K8sGPT, KAIT, and AITERM** represent a significant step towards more autonomous Kubernetes operations.
- ✓ The role of **Kubernetes engineers** is evolving, requiring new skills in  AI fundamentals, data engineering, and strategic thinking.
- ✓ **DevOps Foundations like Cloud, K8s, and Linux** use are crucial as we evolve more into AI-based technologies.
- ✓ Preparing for an **AI-driven future in Kubernetes** requires ongoing learning and adaptation.

© Copyright KodeKloud

Key Takeaways:

1. AI is transforming Kubernetes management, offering improved efficiency, accuracy, and insight.
2. Tools like K8sGPT represent a significant step towards more autonomous Kubernetes operations.
3. The role of Kubernetes engineers is evolving, requiring new skills in AI, data science, and strategic thinking.
4. Ethical considerations and responsible AI use are crucial as we integrate these technologies.

5. Preparing for an AI-driven future in Kubernetes requires ongoing learning and adaptation.

Call to Action



Explore AI-powered
Kubernetes tools



Invest in AI
education



Engage with the
community



Evaluate your
current processes



Develop an AI
strategy

© Copyright KodeKloud

Call to Action:

1. Explore AI-powered Kubernetes tools: Start with K8sGPT or similar tools in a test environment to gain hands-on experience.
2. Invest in AI education: Take online courses or attend workshops to build your AI skills.
3. Engage with the community: Join Kubernetes and AI-focused forums to stay updated and share experiences.

4. Evaluate your current processes: Identify areas in your Kubernetes workflows that could benefit from AI integration.
5. Develop an AI strategy: Begin planning how you'll incorporate AI into your Kubernetes operations over the next 1-3 years.



KodeKloud

© Copyright KodeKloud

Visit www.kodekloud.com to learn more.