

Building Resilience: Effective Backup and Disaster Recovery for Vector Databases on Kubernetes



KubeCon



CloudNativeCon

North America 2024

About Us



Pavan Navarathna
Engineering Manager
Veeam



Shwetha Subramanian
Member of Technical Staff
Veeam

The Rise of AI in Enterprise Applications



Optimized Resource Management

Intelligent Automation

Enhanced Security

Improved User Experience

Predictive Maintenance

Cost Optimization

The Rise of Kubernetes for AI Applications

The background of the slide features a stylized, abstract staircase that ascends from the bottom left towards the top right. The steps are light gray and appear to be floating or made of a translucent material. Behind the staircase, there are several vertical lines of varying heights, creating a sense of depth and structure. Diagonal light rays, in a warm, golden-brown hue, emanate from the base of the staircase and spread outwards, adding a dynamic and futuristic feel to the overall design. The color palette is primarily composed of soft blues, grays, and the warm tones of the light rays.

Faster Deployments

High Availability

Massive Scalability

Reduce Infrastructure Costs

Efficient Backup and Disaster Recovery

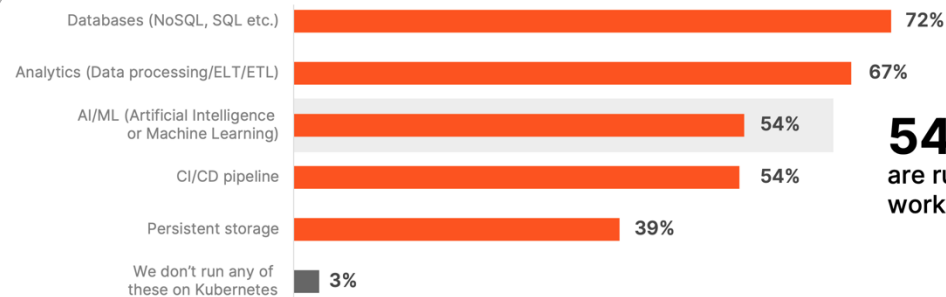
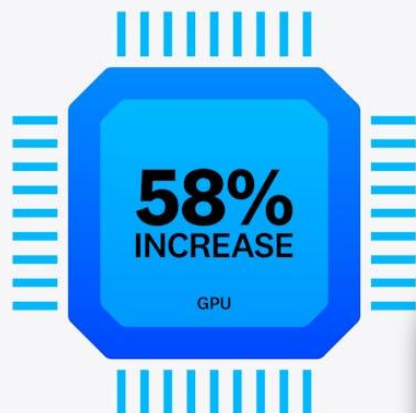
Ecosystem and Community Growth

The Rise of Kubernetes for AI Applications

**CONTAINERIZED
GPU USAGE IS
ON THE RISE**

Source: Datadog

Year-over-year growth of containerized
GPU instance hours



54%
are running AI/ML
workloads on Kubernetes

Source: The Voice of Kubernetes Experts Report 2024

Introduction to Vector Databases

Essential for High-Dimensional Data Storage

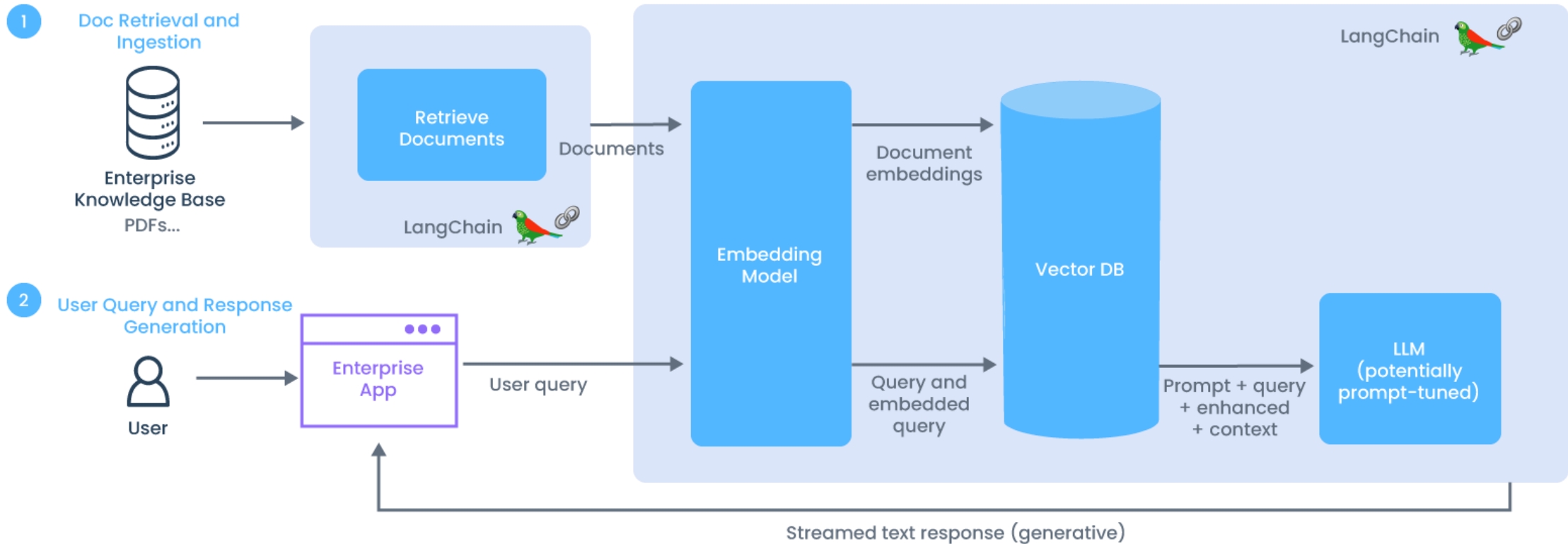
Core to Retrieval Augmented Generation (RAG)

Enhancing Contextual Relevance in AI Responses

Vector databases store high-dimensional vectors representing unstructured data like text, images, and audio.

Vector Databases in RAG

RETRIEVAL AUGMENTED GENERATION (RAG) SEQUENCE DIAGRAM



Why Vector Databases Need Data Protection

Data Integrity and Security

Disaster Recovery

Cost and Time Efficiency

Version Control

Regulatory Compliance

“

I'm Vector, I commit
crime with both
Direction and
Magnitude

”





Introduction to Kanister

CNCF Sandbox – September 2023

Kanister Controller

Operator responsible for state management of Kubernetes Custom Resources

Blueprints

Define workflows for backup, restore and delete operations

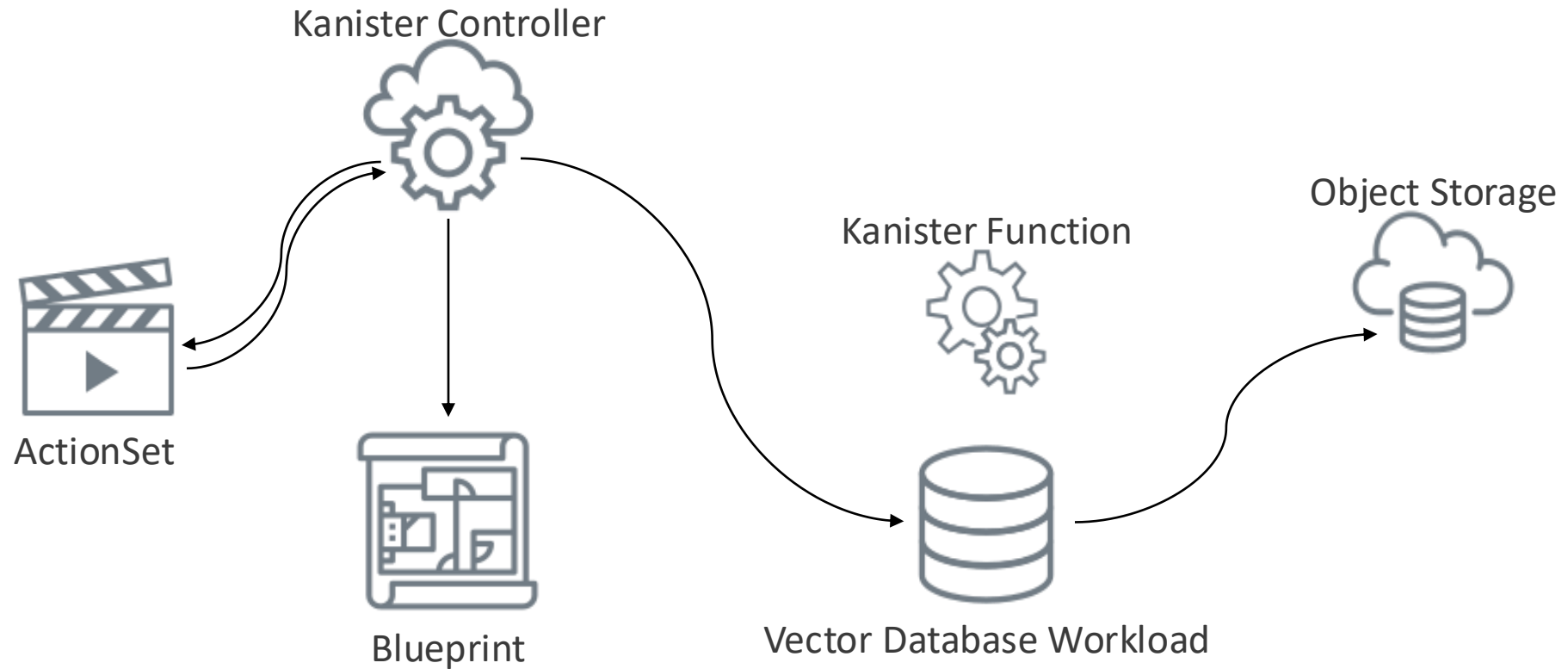
ActionSets

Run an action from a blueprint

Profiles

Define target destination for backups or a source for restores

How Kanister Protects Vector Databases



Demo Overview

BookNest: Book Recommendation Chatbot

*Question-answer application: semantic vector search
on a vector database of books*

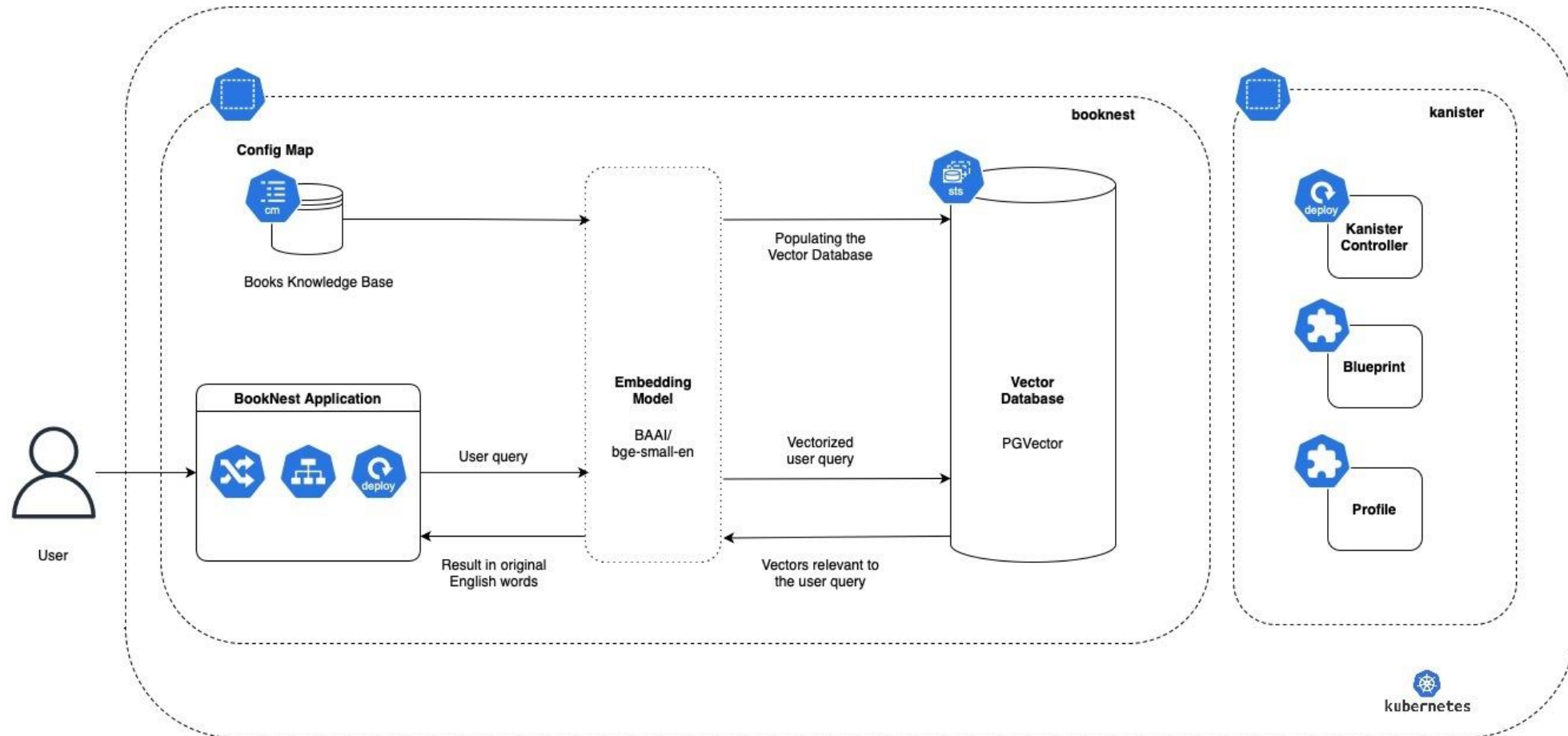
384 Dimension Vector

Bitnami PostgreSQL with PGVector Extension

*BAAI (Beijing Academy of Artificial Intelligence)
General Embedding Model*



Demo Architecture



DEMO



Your Next Steps

Protect Your Data on Kubernetes

Join the Kanister Community

Visit Us at the Kanister Project Pavilion
Booth or the Veeam Booth (K7)

Resources

[Kanister](#)

[Bitnami Postgresql](#)

[PGVector](#)

[Backup and Recovery for Vector Databases on Kubernetes Using Kanister](#)

[The Voice of Kubernetes Experts Report 2024](#)

[Datadog Report](#)

