

Red Hat
Summit

Connect

Come sfruttare l'implementazione dell'AI per ottimizzare lo sviluppo infrastrutturale e semplificare il processo di troubleshooting

Maurizio Romani, Solutions Architect at Elastic

Luigi Ria, Director of Professional Services at Kiratech

Milano, 19/11/2024

 **KIRATECH**[®]
PLATFORM HEROES

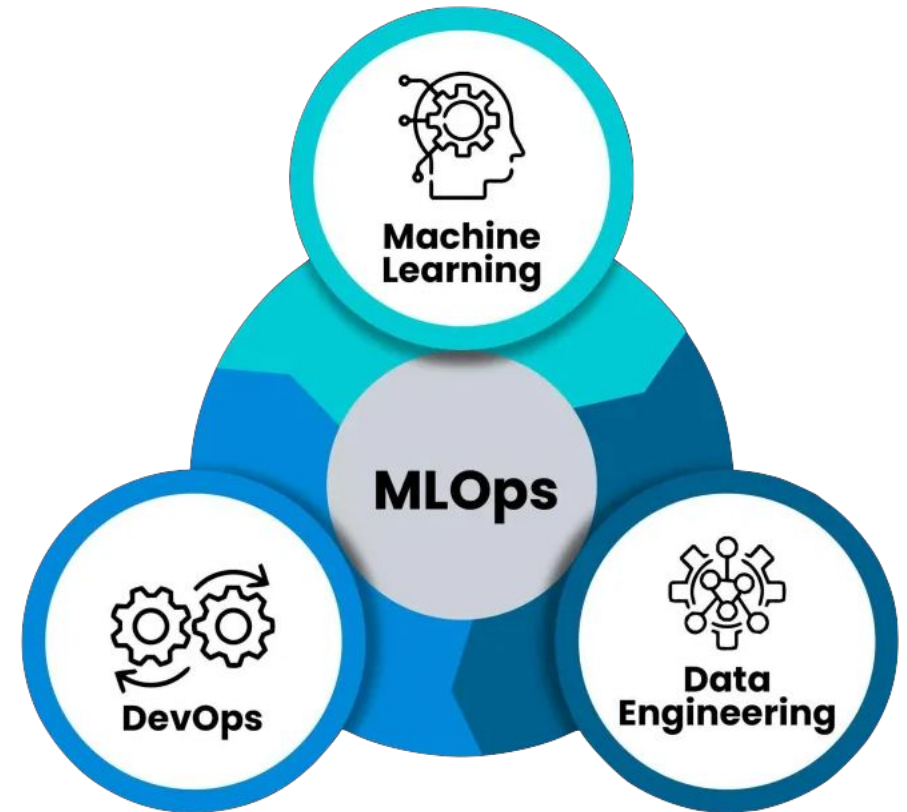
 elastic

 **Red Hat**

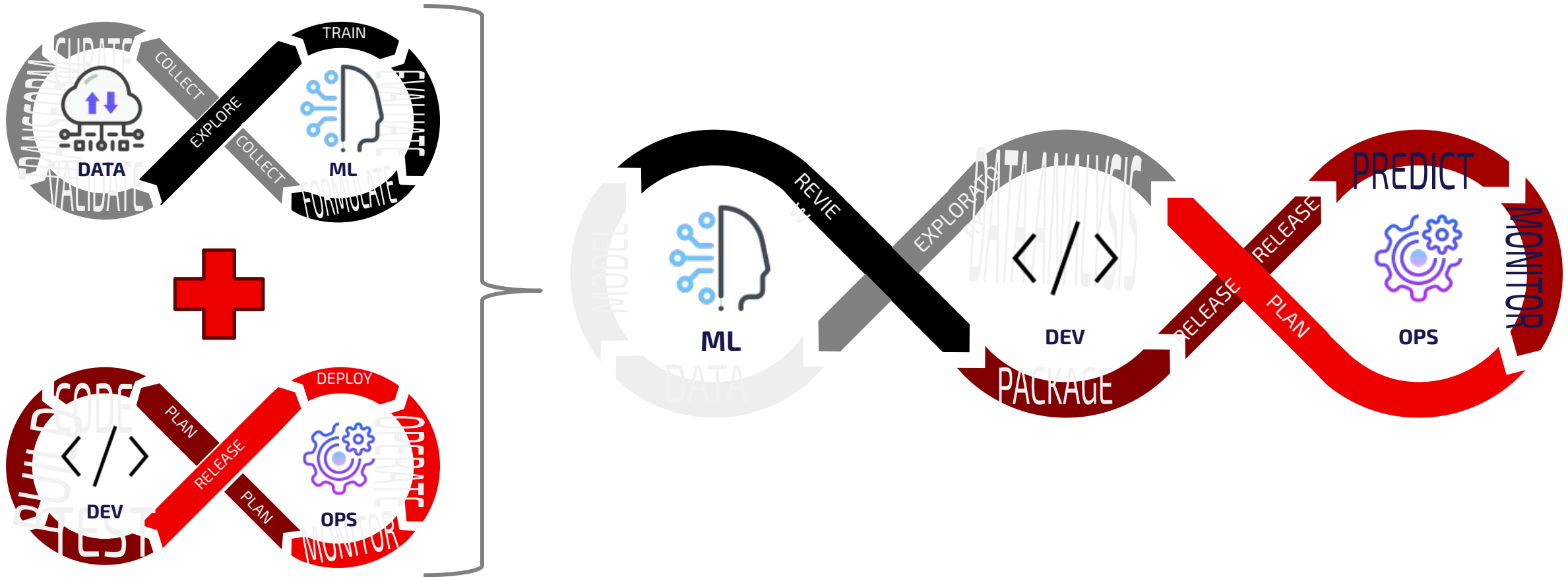
MLOPS | Why do we need MLOps

The better question would be to ask “**Why not?**”. When MLOps best practices are integrated into the business model, a plethora of advantages become **clear and the implementation of data-centric AI becomes easier.**

As with anything in the business world, a cost-benefit analysis is necessary and it's up to you to decide what works best.



DEVOPS + ML = MLOPS

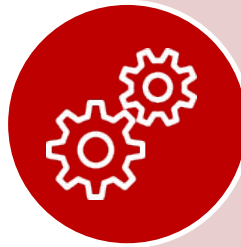


COLLABORATION AND CODE VERSIONING



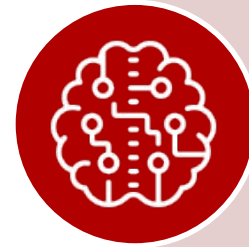
- Versioning
- Collaboration
- Code Automation

TRAINING AUTOMATION



- CI/CT/CD
- Model Training

MODEL MONITORING AND MANAGEMENT



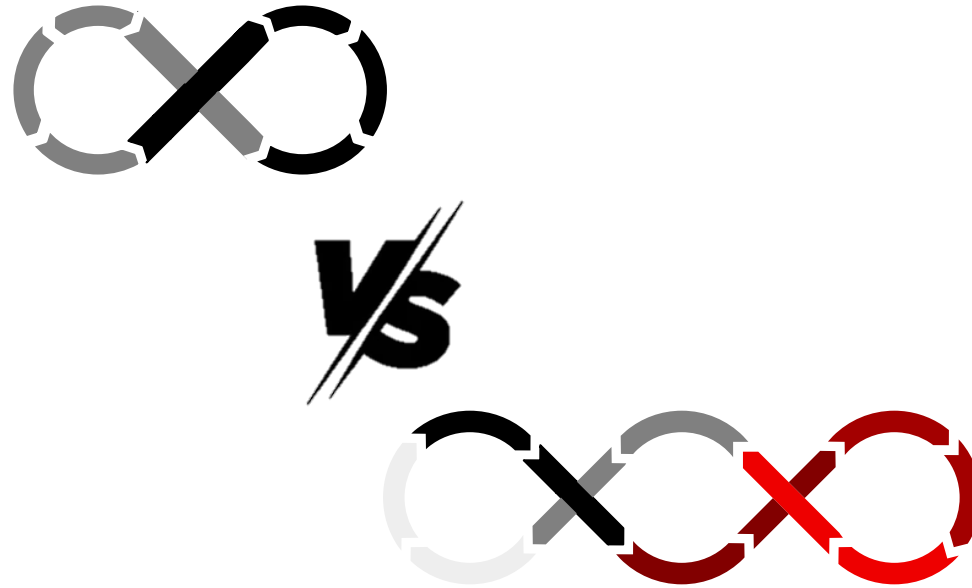
- Model lifecycle
- Model Training
- Model Quality
- Monitoring (Pipeline & Model)

SECURITY AND GOVERNANCE

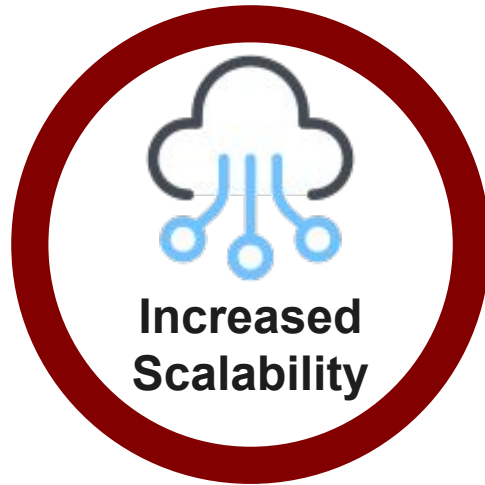


- Policy management
- Data monitoring
- Experiment monitoring

While **DevOps** focuses on software development and release pipelines, MLOps extends this approach to AI models by integrating datasets, machine learning experiments, and models into the lifecycle.



MLOps integrates with platform engineering by creating an ecosystem where data scientists, DevOps engineers, and developers collaborate, enabling automation of AI infrastructures and processes.



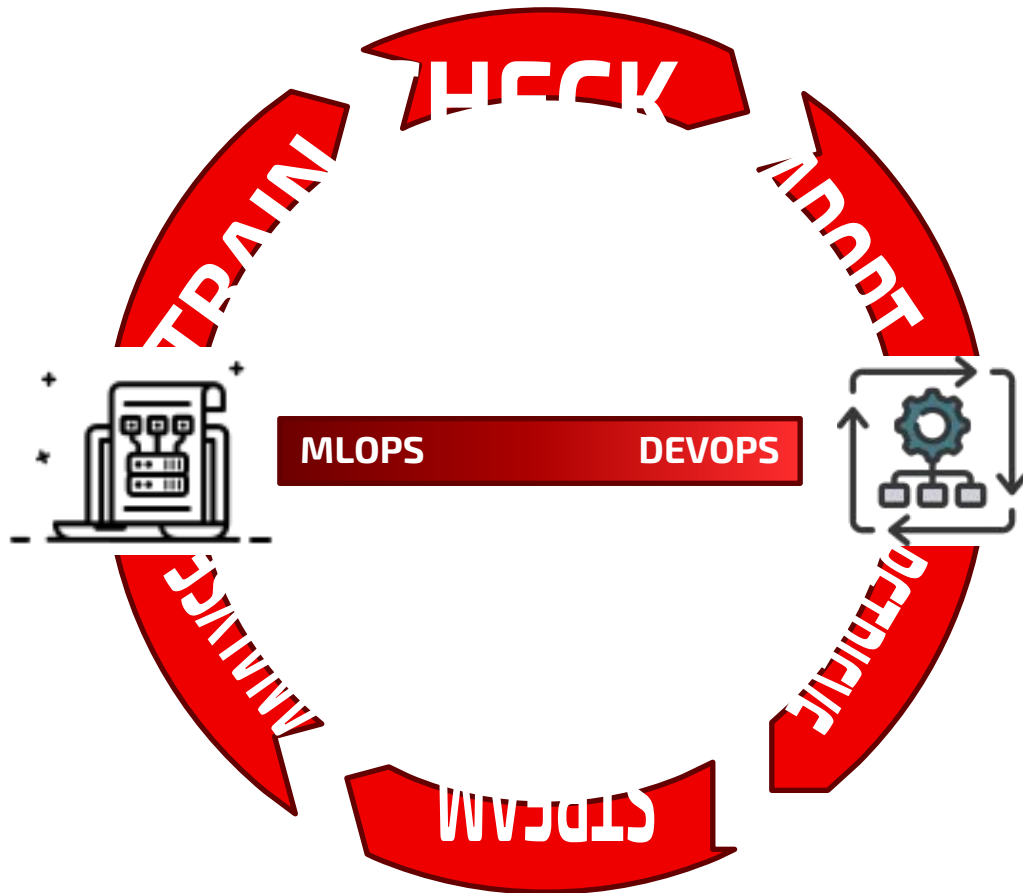
MLOPS allows you to share business targets by integrating them into the evolution of the model. This approach allows you to quickly scale model features and transform them into ROIs



A primary element for security is the **monitoring of the evolution of the model.** Monitoring starts with tools integrated within the toolchain that control and prevent model drift, ensuring its reliability



The use of the MLOps toolchain allows you to improve the models to be applied in **FinOps processes** to improve the infrastructure investment with constant performance



The shift from MLOps to PlatformOps allows infrastructure to be dynamically managed based on model-driven decisions. This approach, enhanced by DevSecOps integration, enables:

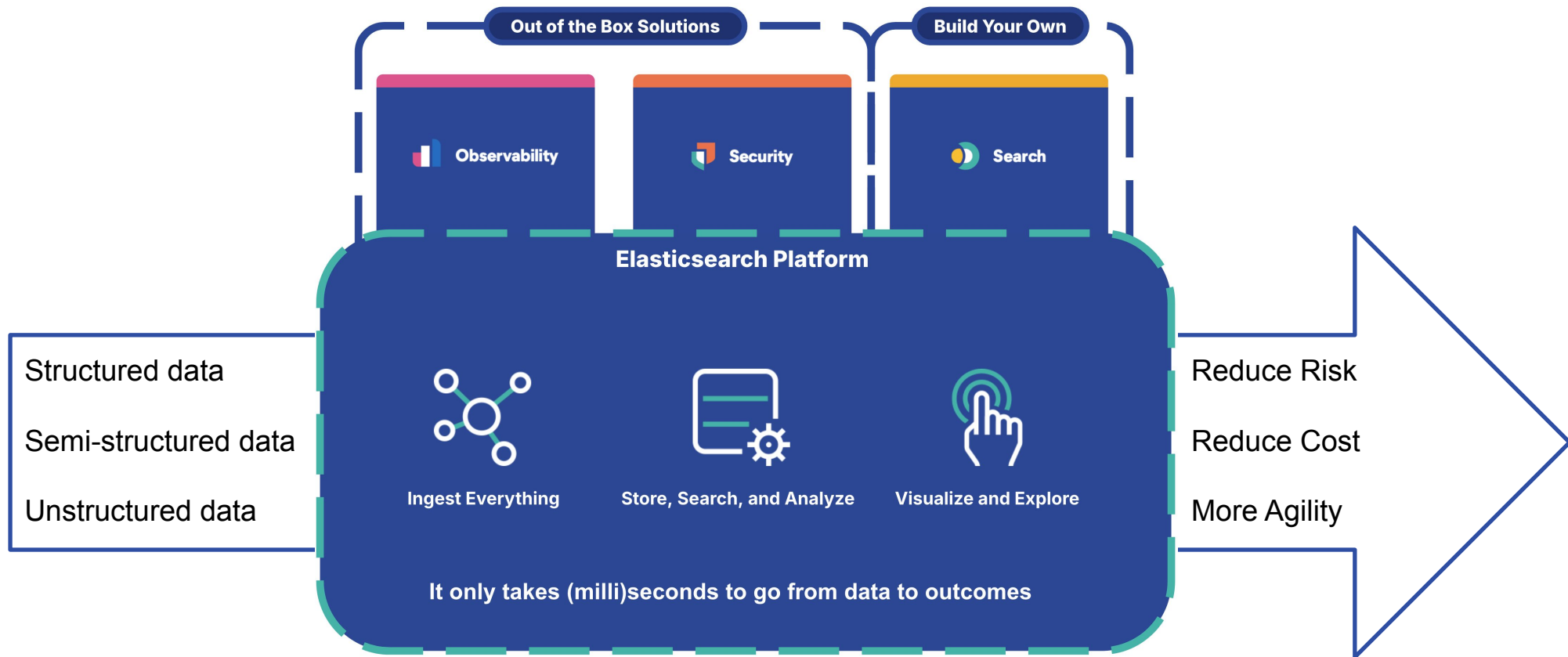
- ❑ **Dynamic Adaptation:** Infrastructure resources and security adjust in real time based on model outputs.
- ❑ **Managed Orchestration:** The infrastructure orchestrator, driven by AI models, allocates resources, scales containers, and updates data pipelines.
- ❑ **Flexibility and Security:** Infrastructure changes occur with continuous security monitoring, ensuring operational stability.

This evolution supports a continuous loop of optimization and security, aligned with the dynamic demands of incoming data.

The Elastic Search AI Platform

Meet the Elasticsearch platform:

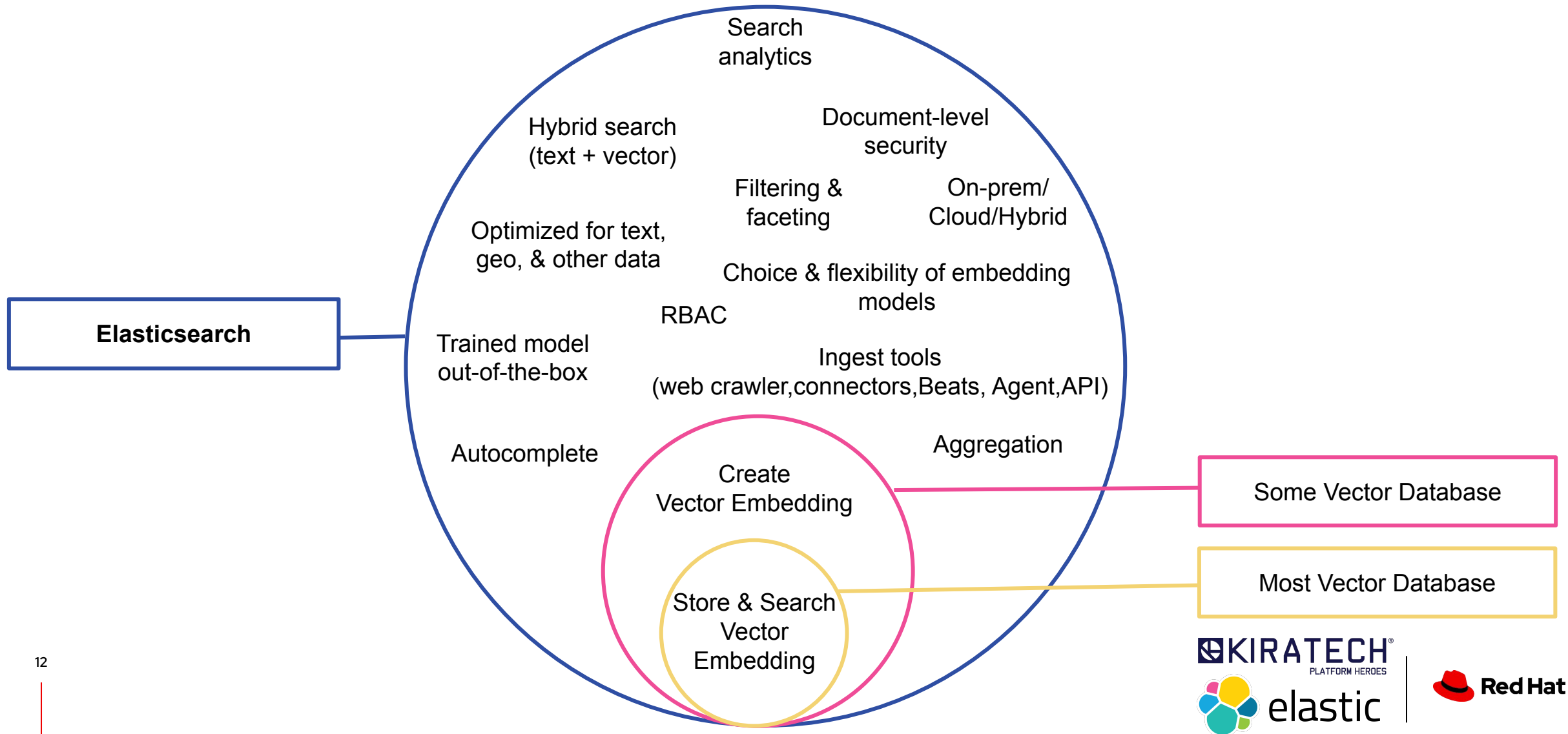
A **unified** data platform to accelerate your **mission**



We've Come a Long Way

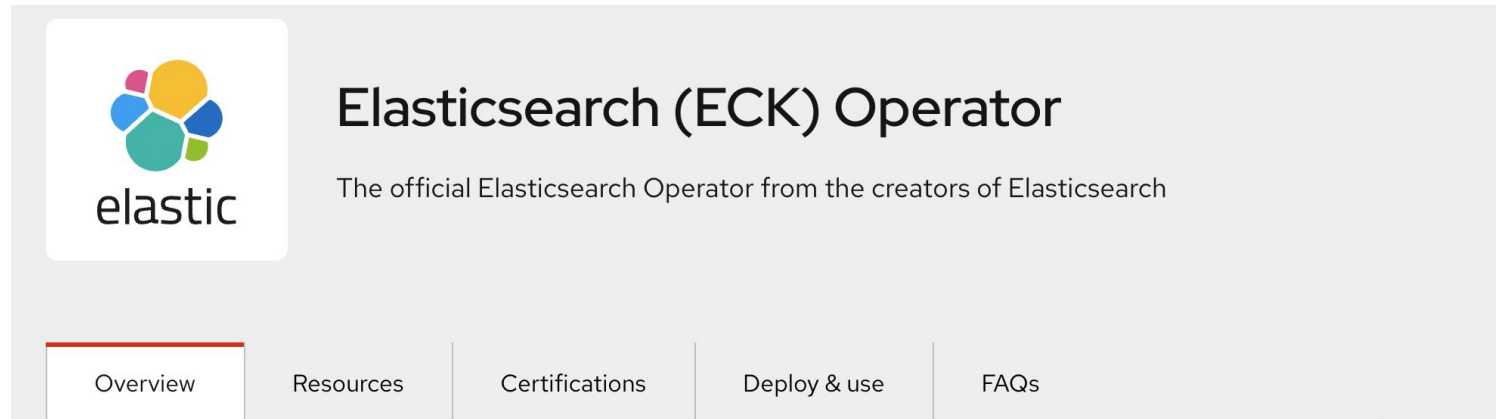



Elastic provides all the capabilities you need in GenAI



Elastic and OpenShift

Elastic Cloud on Kubernetes (ECK) is a Red Hat OpenShift Certified Operator




elastic

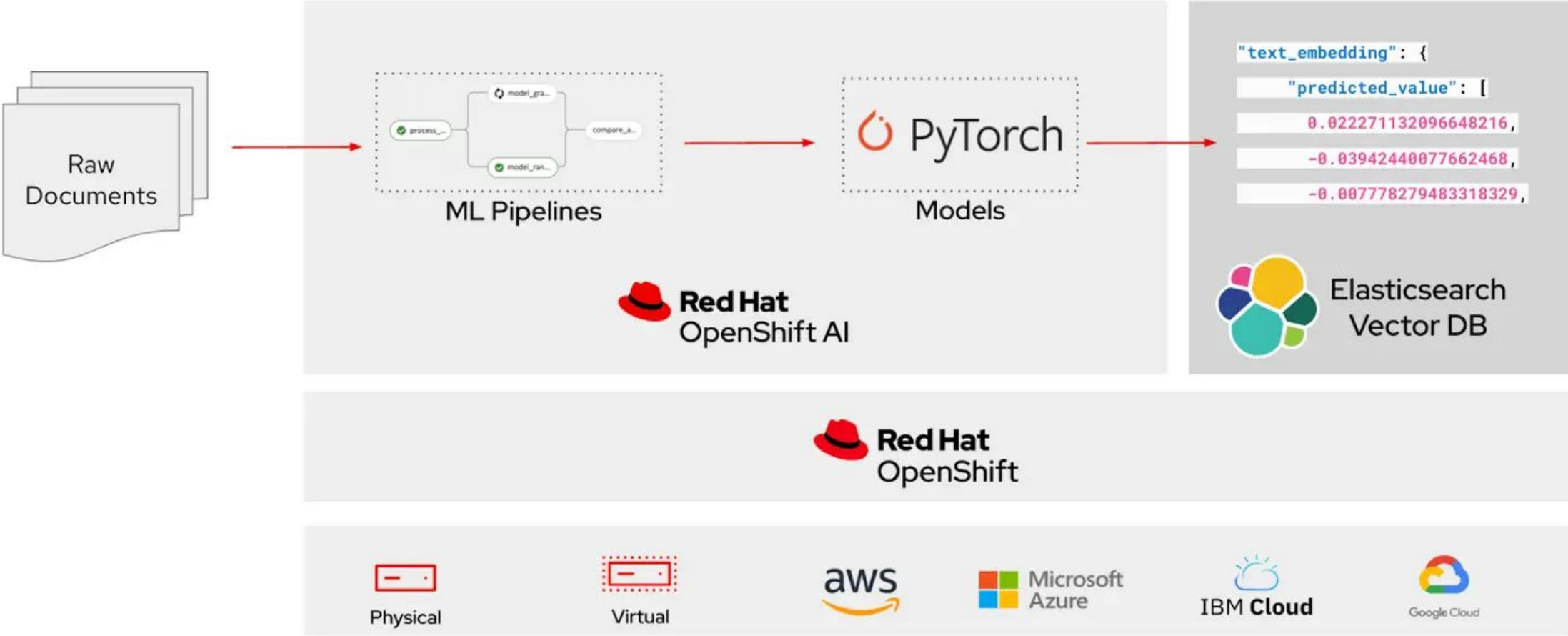
Elasticsearch (ECK) Operator

The official Elasticsearch Operator from the creators of Elasticsearch

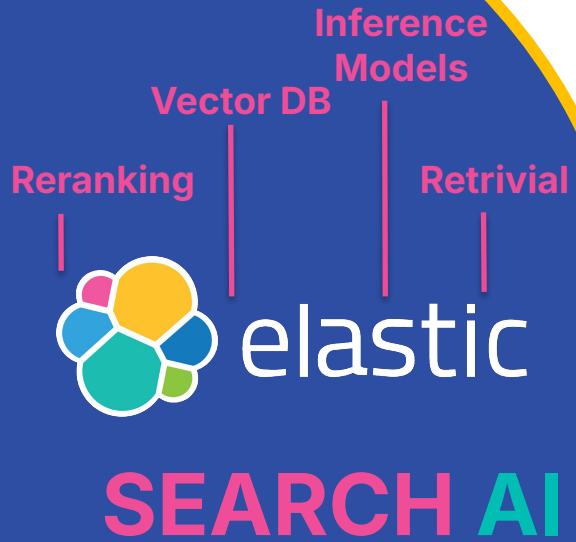
- Overview
- Resources
- Certifications
- Deploy & use
- FAQs

Elastic Cloud on Kubernetes automates the deployment, provisioning, management, and orchestration of Elastic Stack (e.g. Elasticsearch and Kibana) on Kubernetes. The Operator brings the power of Elastic Enterprise Search, Observability, and Security to Kubernetes. We offer both a Basic and Enterprise license with the Operator. The Basic license is free and comes with a host of features like multi-cluster management. The Enterprise license is paid and great for customers who want 24/7 support and access to advanced features such as cross-cluster search and cross-cluster replication.

Red Hat OpenShift AI & Elastic Vector Database



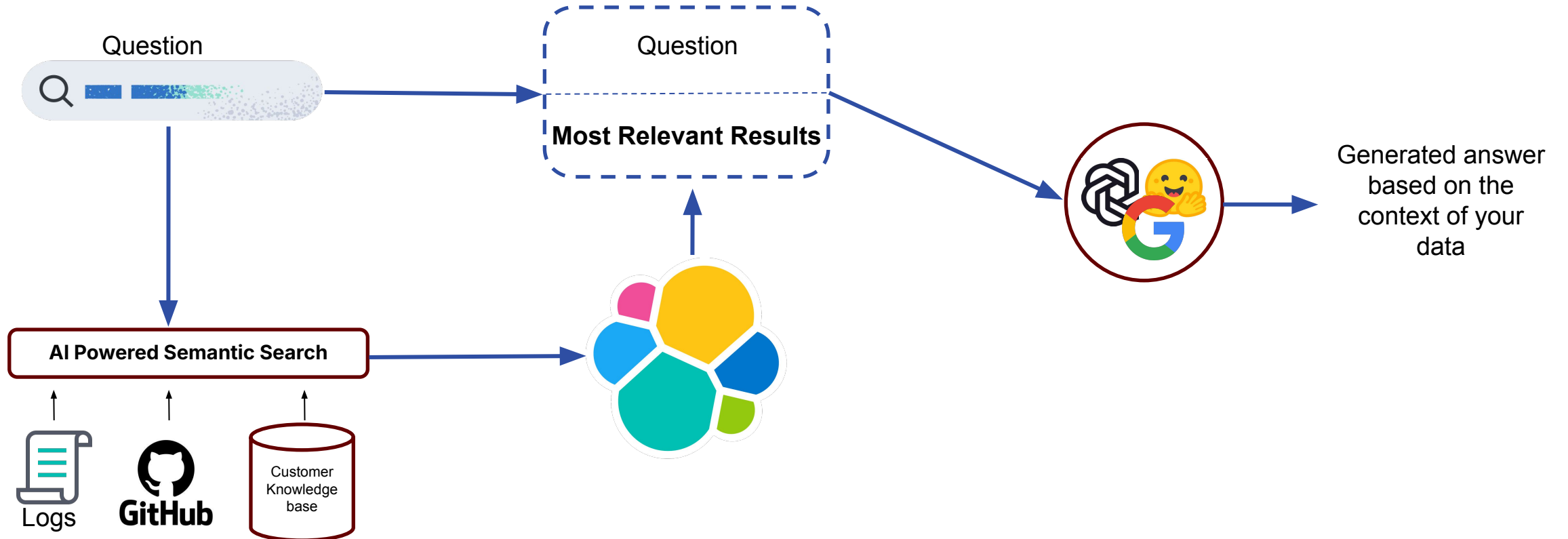
Enabling Technologies



Solutions Enabled



Retrieval Augmented Generation with Elastic



The Elastic AI Assistant for Observability



Contextual Prompts



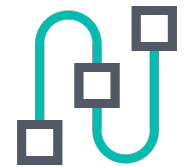
Empower your Engineers



Combine with Internal KB



Open and Transparent








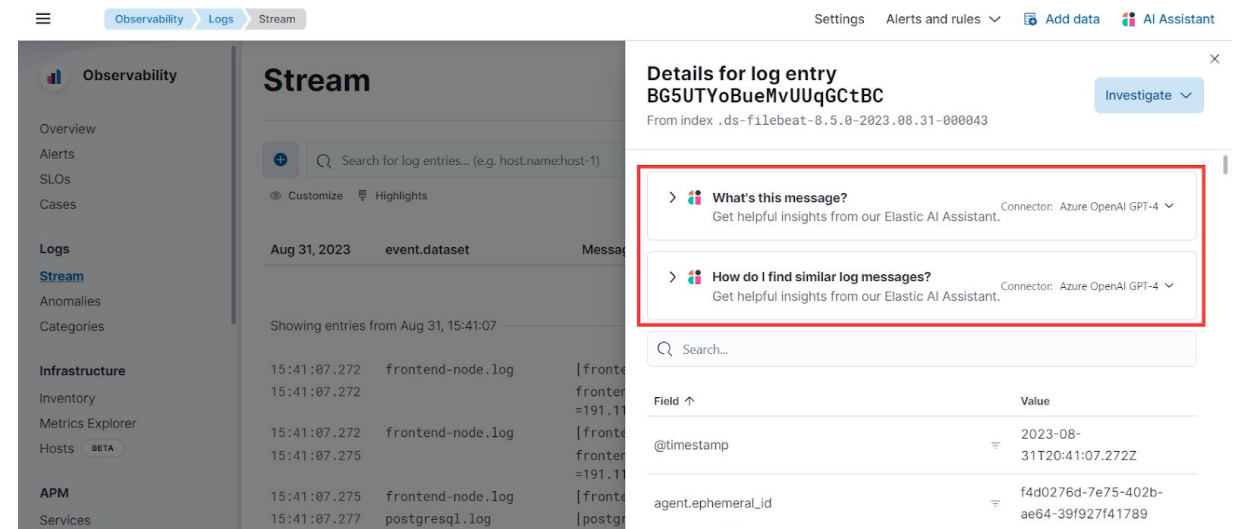
Multi-Model Support

For Every Observability Engineer The Elastic AI Assistant



AI Assistant Observability in context

-  Provide Alerting RCA
-  Explain log messages
-  Explain host processes
-  Explain APM errors
-  Profiling to analyze libraries and functions



AI Assistant in Observability use cases



Explain and Solve

Act as an internet search tool to explain error codes, alert conditions, and provide a natural language chat interface to provide solutions.



AI for Elastic expertise

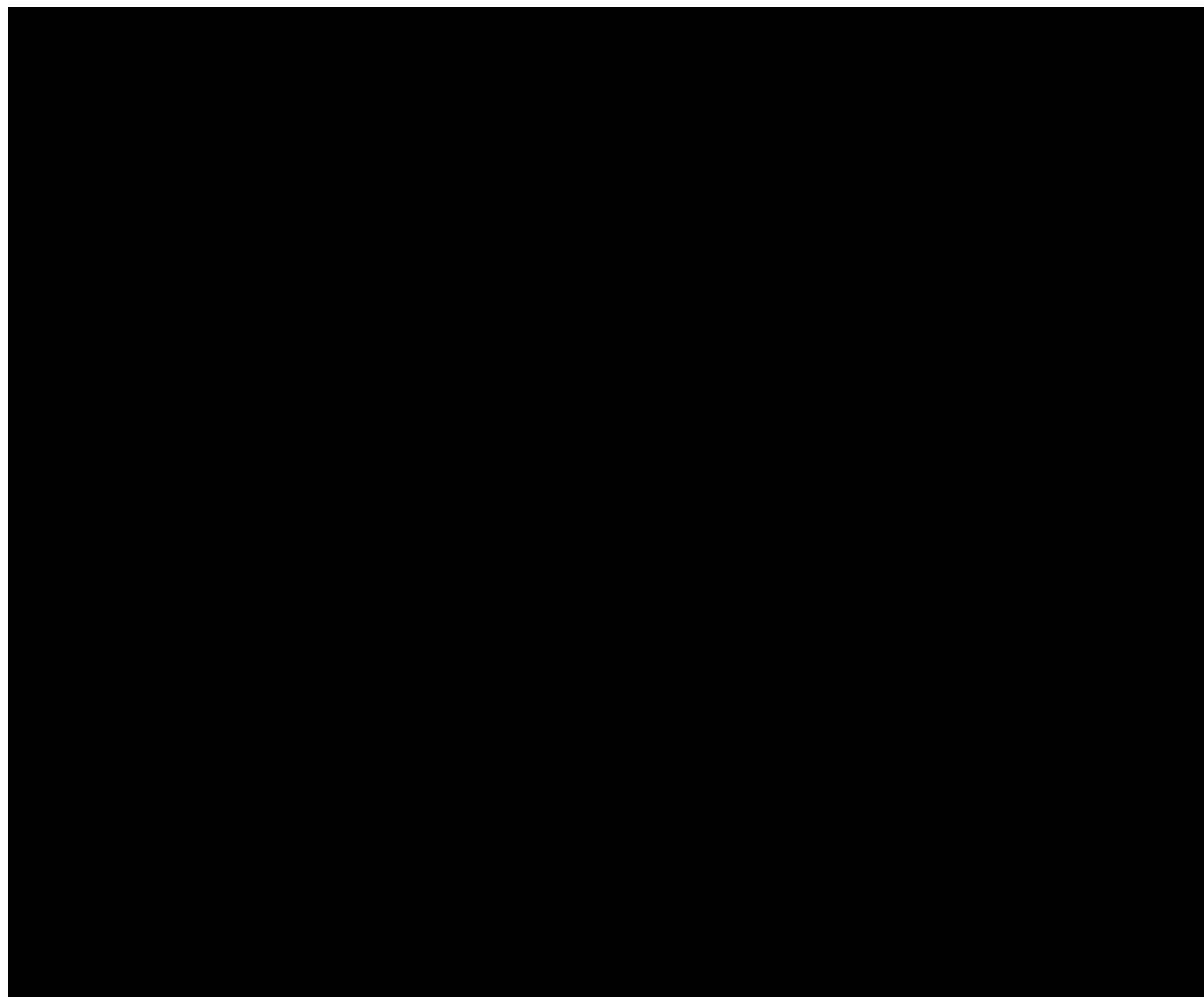
Not sure how to create that query or build the visualization? Need to translate a query or synthetic monitor? We have a natural language UI that will build it for you and provide the code or syntax to execute on what you need



Unlock internal knowledge

Ingest internal knowledge, such as the content of private runbooks written by our SREs, github issues, post-mortem reports, internal mailing lists etc. Query the private data to find information that may be relevant in answering the particular query that we are about to send to the LLM.

AI Assistant in Action



Red Hat
Summit

Connect

Q&A



Red Hat
Summit

Connect

Thank you

