

BeamStack: An Open source Framework for running Machine Learning Pipelines with Apache Beam

Olufunbi Babalola & Mat Fait



BEAM
S U M M I T

September 4-5, 2024

Sunnyvale, CA. USA

About the Presenters



Olufunbi Babalola is a Technical Product Manager at MavenCode. He has an extensive background in cloud and data engineering, with a focus on developing new innovative Artificial Intelligence and Machine Learning products. He has over 4 years of experience conceptualizing and delivering large-scale applications. Olufunbi holds a Master's degree in Software Management from the Integrated Innovation Institute, Carnegie Mellon University.



Mathew Fait is an experienced Product Manager at MavenCode, responsible for managing daily developer operations. With a strong background in product management, he focuses on streamlining workflows, fostering innovation, and ensuring product development aligns with business objectives.



About MavenCode

MavenCode is an Artificial Intelligence Solutions Company with HQ in Dallas, Texas and a remote delivery workforce across multiple time zones. We do training, product development and consulting services with specializations in:

- Provisioning Scalable AI Ops, ML Ops and LLM Ops Infrastructure - OnPrem and In the Cloud
- Development & Production Operationalization of ML platforms - OnPrem and In the Cloud
- Streaming Data Analytics and Edge IoT Model Deployment for Federated Learning
- Building out Data lake, Vector Stores, Feature Store, and ML Model Management platform



twitter.com/mavencode



BEAM
SUMMIT

Agenda

- Overview of Beamstack
- Why do you need Beamstack?
- How to use Beamstack
- Product Roadmap
- Call for Participation
- Questions & Answers



Overview of Beamstack



BEAM
SUMMIT

What is Beamstack?



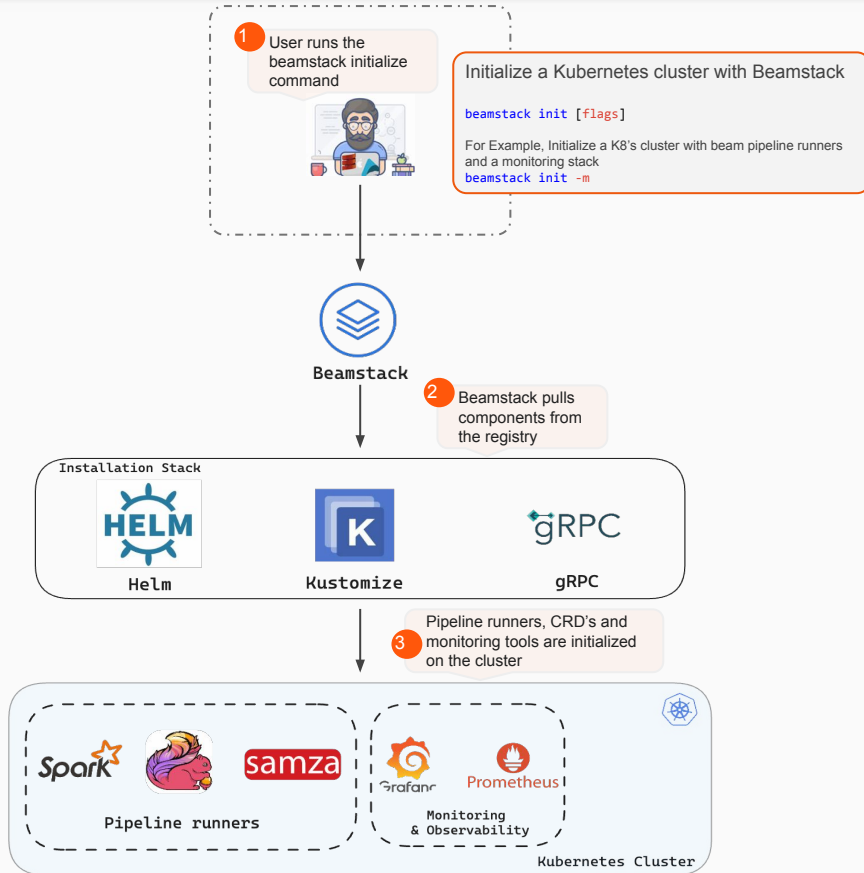
Beamstack

- Beamstack is a framework that makes it easy to run infrastructure agnostic low-code Gen AI and ML workflows
- Beamstack comes with a robust CLI that abstracts most of the complex Ops around Gen AI and ML workflows deployments



BEAM
SUMMIT

How does Beamstack work?



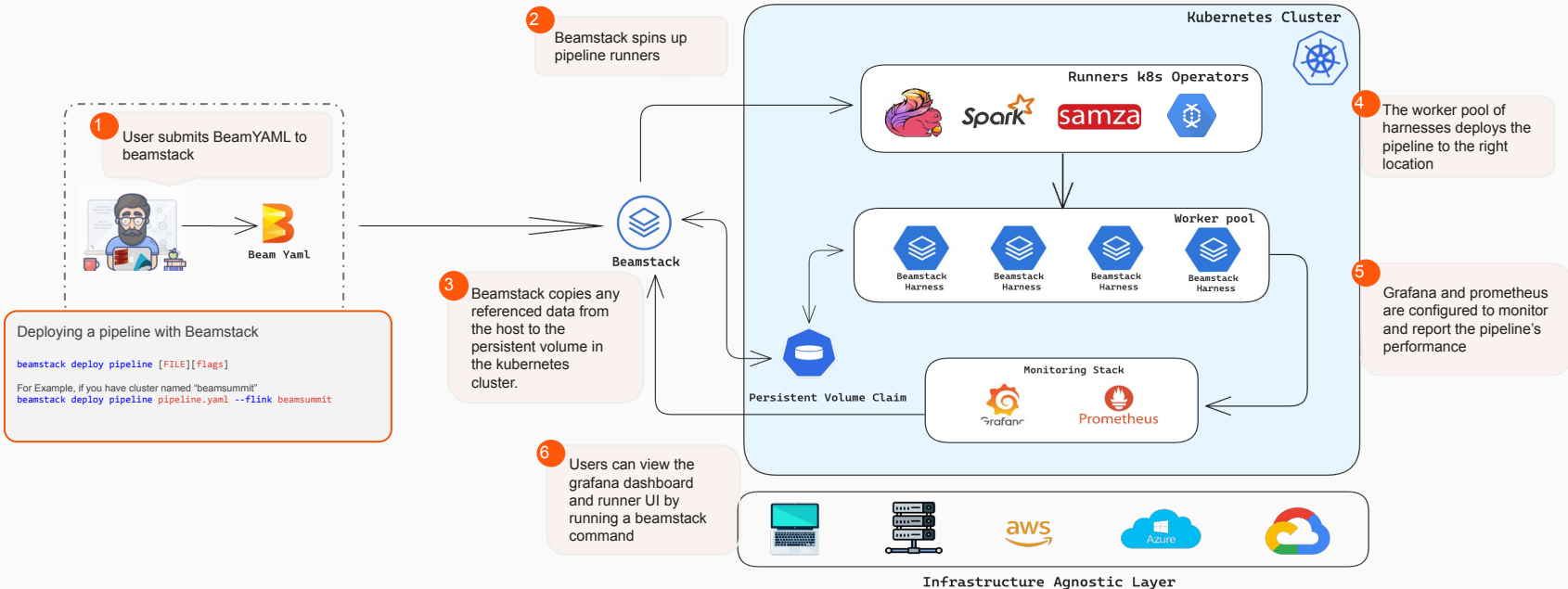
User runs the `beamstack init -m` command in the beamstack CLI.

Beamstack pulls the cluster components to be installed from the registry using helm package manager and kustomize.

The kubernetes cluster is initialized with beam pipeline runners(flink, spark or samza) and a monitoring stack(Grafana & Prometheus).



Beam YAML makes this easy!



Why do you need Beamstack ?



BEAM
SUMMIT

Why Beamstack?



Configurable Deployment Environment

With minimal steps you can setup Beamstack on your local minikube, bare metal or cloud infrastructure



Ease of Deployment with Beam Low-code YAML

Beamstack adopts a low-code approach towards pipeline deployment which makes the process easier and faster



Composable and Reusable Pipeline Components

Reusable pipeline components designed for easy composition and customization, enabling efficient workflow creation and deployment.

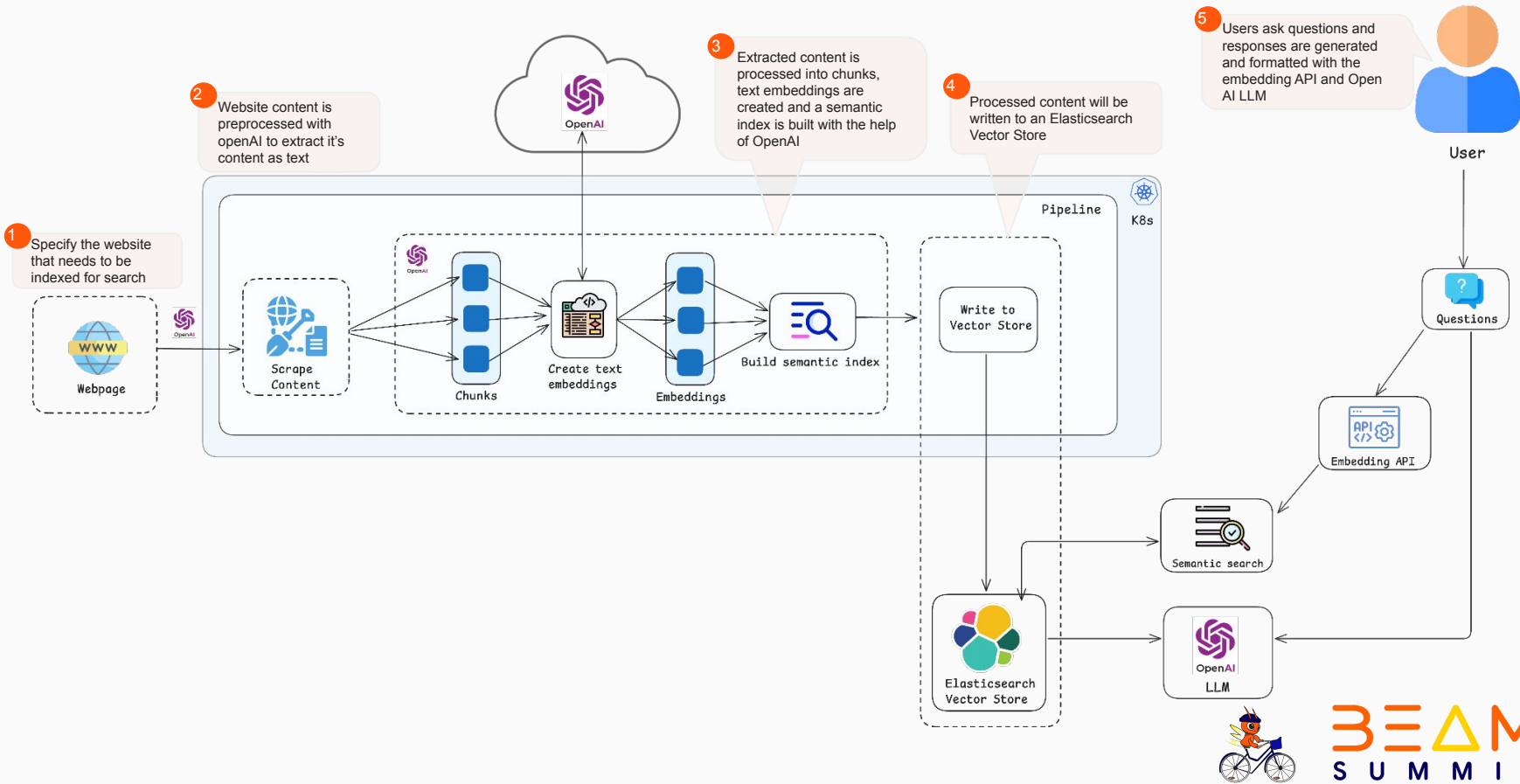


Collaborative Setup for Development Teams

Beamstack's modular architecture facilitates collaborative setup's for various technical teams within an organization



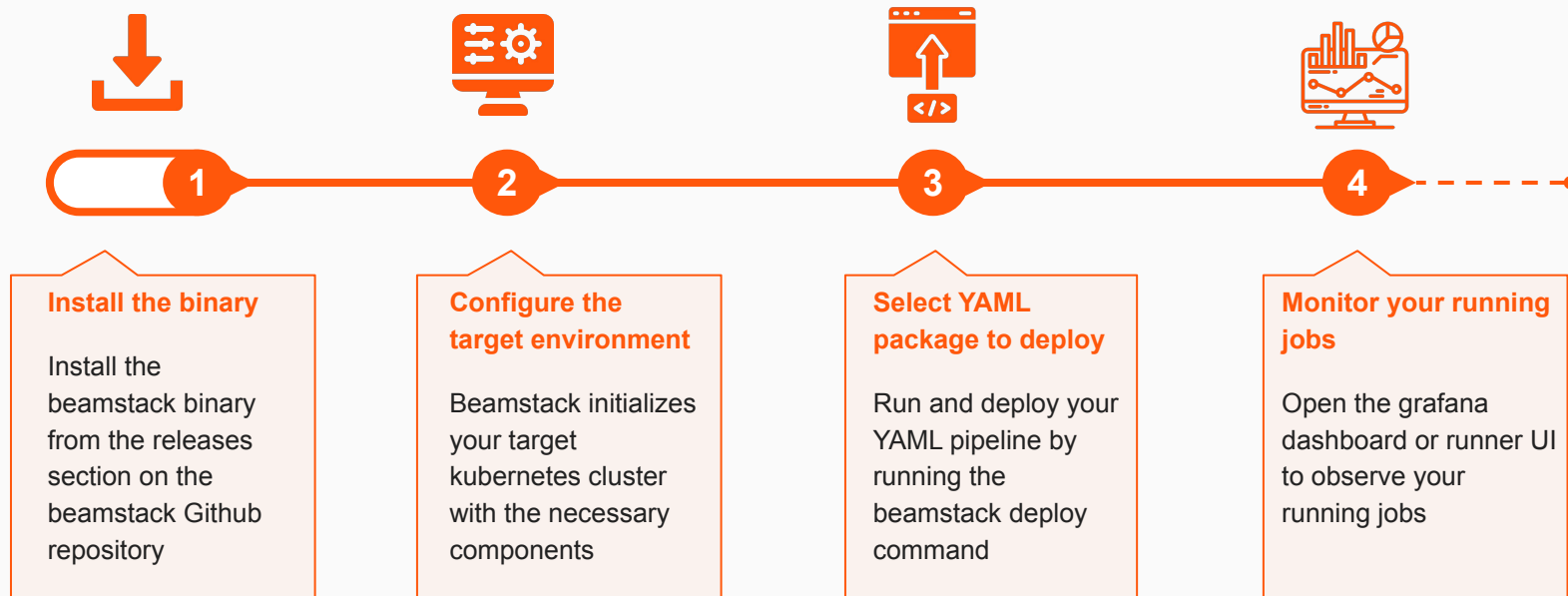
Example Use Case: Creating Text Embedding + Saving it to Vector Database



How to use Beamstack



How to use Beamstack

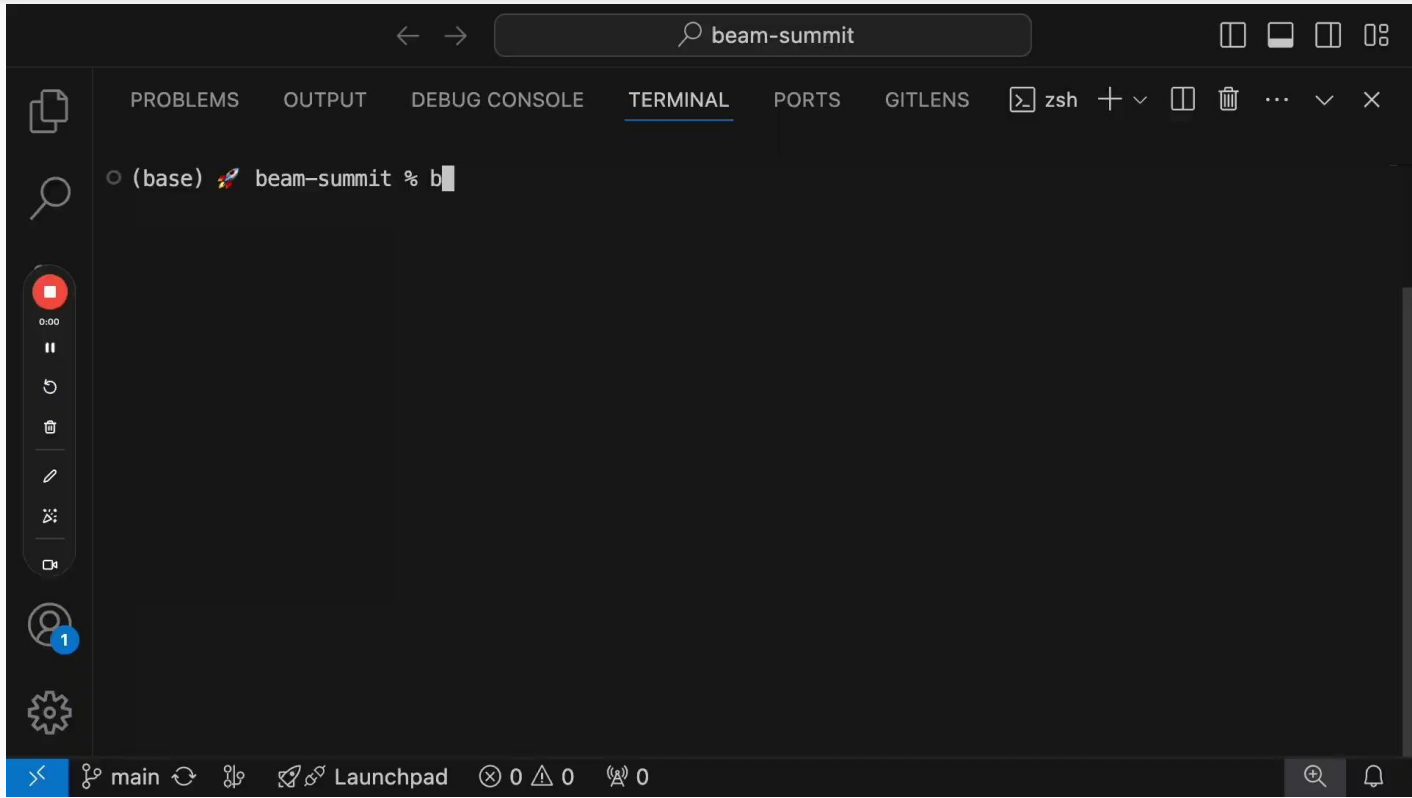


Configure the target environment



Configure the target environment

Beamstack initializes your target kubernetes cluster with the necessary components

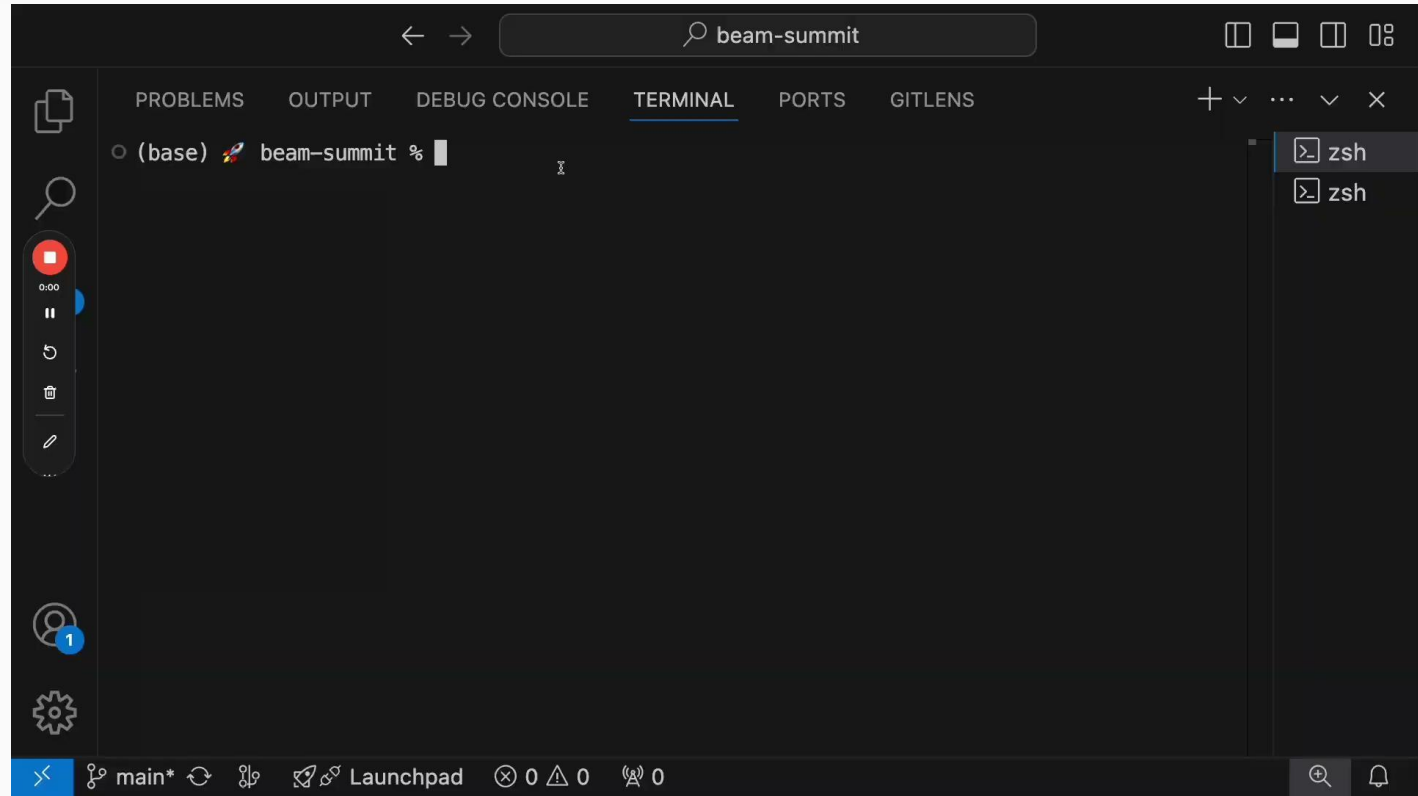


Select YAML pipeline to deploy



Select YAML package to deploy

Run and deploy your YAML pipeline by running the beamstack deploy command

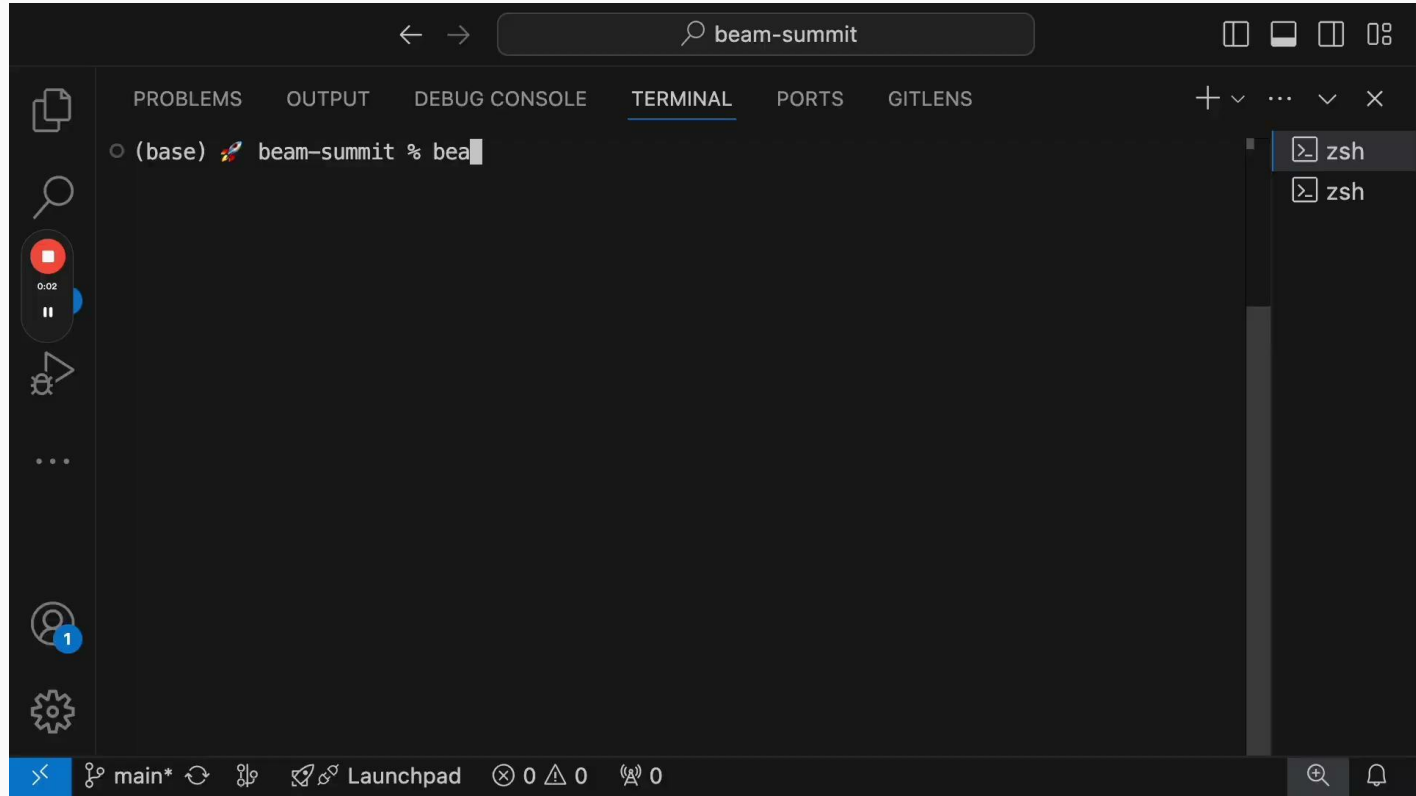


Monitor your running jobs



Monitor your running jobs

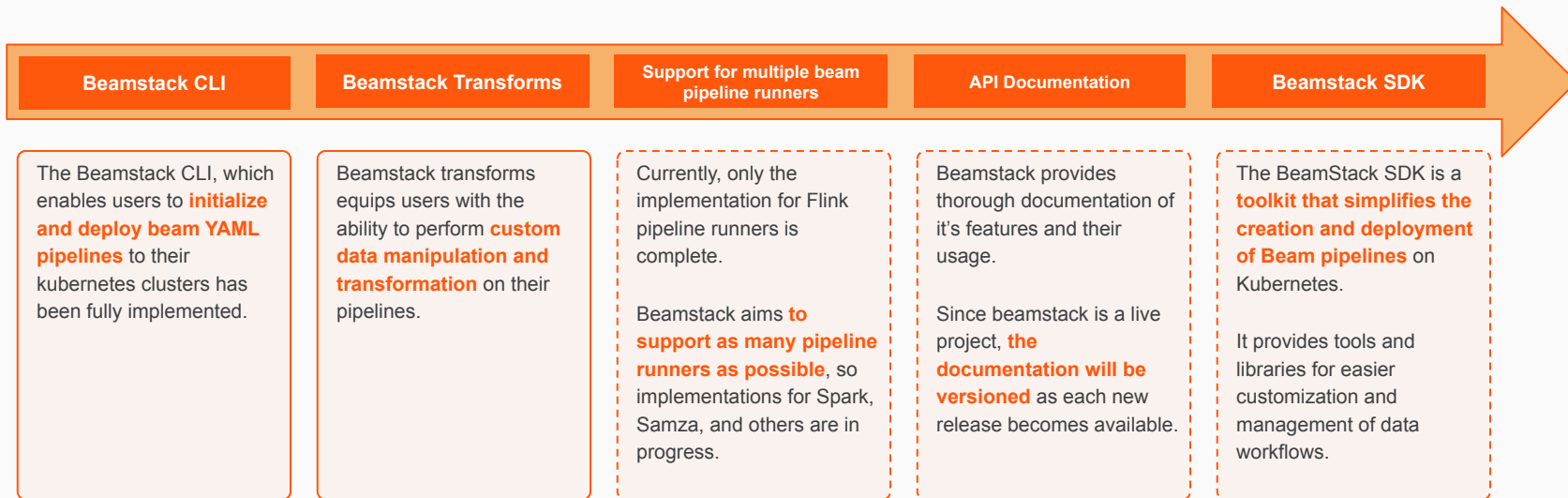
Open the grafana dashboard or runner UI to observe your running jobs



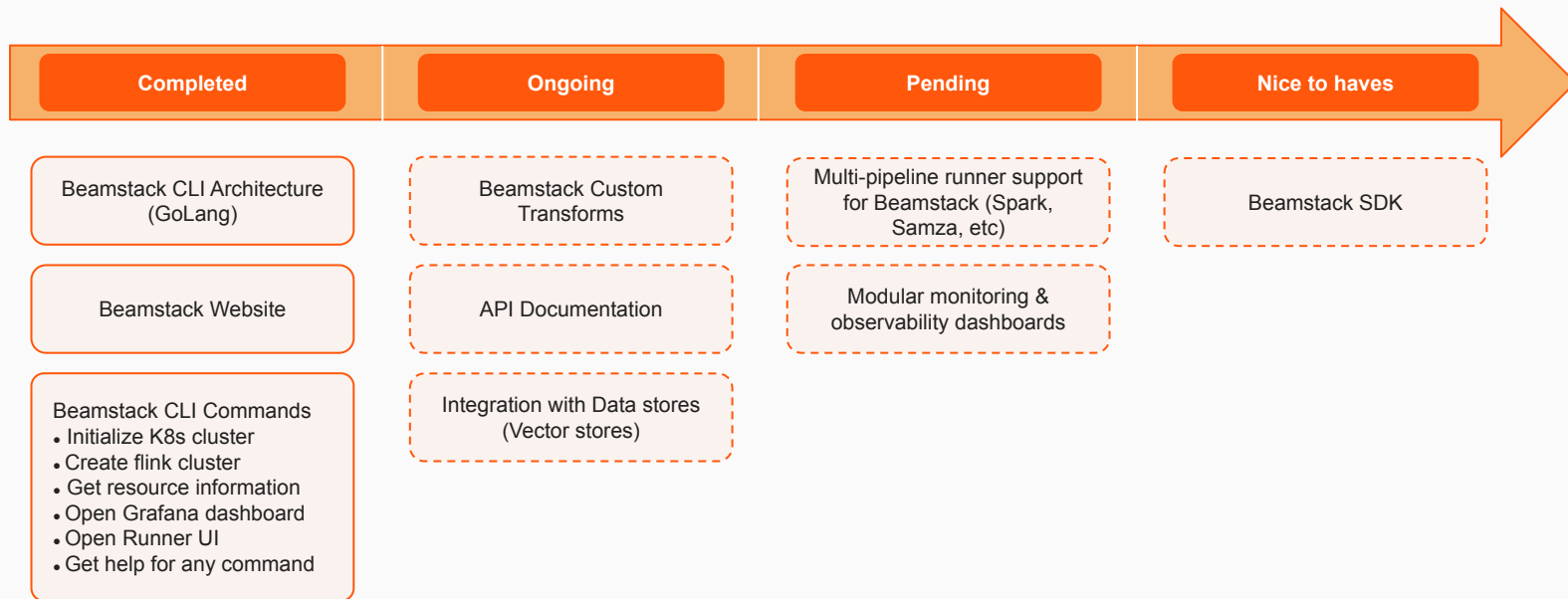
Beamstack Roadmap



Beamstack Roadmap: Releases



Beamstack Roadmap: Tasks breakdown



Beamstack Roadmap: Feature Tracker

BeamStackProj / Projects / Beamstack

Beamstack

Backlog Team capacity Current iteration Roadmap My items + New view

Filter by keyword or by field

- Todo** 5 / 5 Estimate: 0
This item hasn't been started
 - beamstack-cli #31 Package the SDK
 - beamstack-cli #32 Implement Proper Error Handling
 - beamstack-cli #33 Develop Convenience Methods
 - beamstack-cli #34 Implement CI/CD
 - beamstack-cli #35 Maintain and Update
- In Progress** 4 / 10 Estimate: 0
This is actively being worked on
 - beamstack-cli #27 API Documentation
 - beamstack-cli #28 Integration with Data stores
 - beamstack-cli #29 Multi-pipeline runner support for Beamstack
 - beamstack-cli #30 Modular monitoring & observability dashboards
- QA** 1 Estimate: 0
 - beamstack-cli #26 Beamstack Custom Transforms
- Done** 1 Estimate: 0
This has been completed
 - beamstack-cli #10 Implement deployment commands for various Beamflow resources.
Iteration 6

Feature Tracker on Github is how we track features to be implemented. We are actively seeking contributors for these new features.



Beamstack Roadmap: Feature Tracker

The screenshot shows a GitHub Project board for the 'Beamstack' project. The board is organized into four columns: 'Todo' (6 items), 'In Progress' (4 items), 'QA' (1 item), and 'Done' (1 item). Each item is represented by a card with a title and a status indicator. The 'QA' column contains a card for 'beamstack-cli #28' titled 'Beamstack Custom Transforms'. The 'Done' column contains a card for 'beamstack-cli #10' titled 'Implement deployment commands for various Beamflow resources'. The board also includes a search bar, a filter dropdown, and a 'Filter by keyword or by field' input. A recording overlay is visible on the right side of the screen.

Column	Item ID	Item Title
Todo (6/5)	beamstack-cli #31	Package the SDK
	beamstack-cli #32	Implement Proper Error Handling
	beamstack-cli #33	Develop Convenience Methods
	beamstack-cli #34	Implement CI/CD
	beamstack-cli #35	Maintain and Update
	In Progress (4/10)	beamstack-cli #27
beamstack-cli #28		Integration with Data stores
beamstack-cli #29		Multi-pipeline runner support for Beamstack
beamstack-cli #30		Modular monitoring & observability dashboards
QA (1)	beamstack-cli #28	Beamstack Custom Transforms
Done (1)	beamstack-cli #10	Implement deployment commands for various Beamflow resources

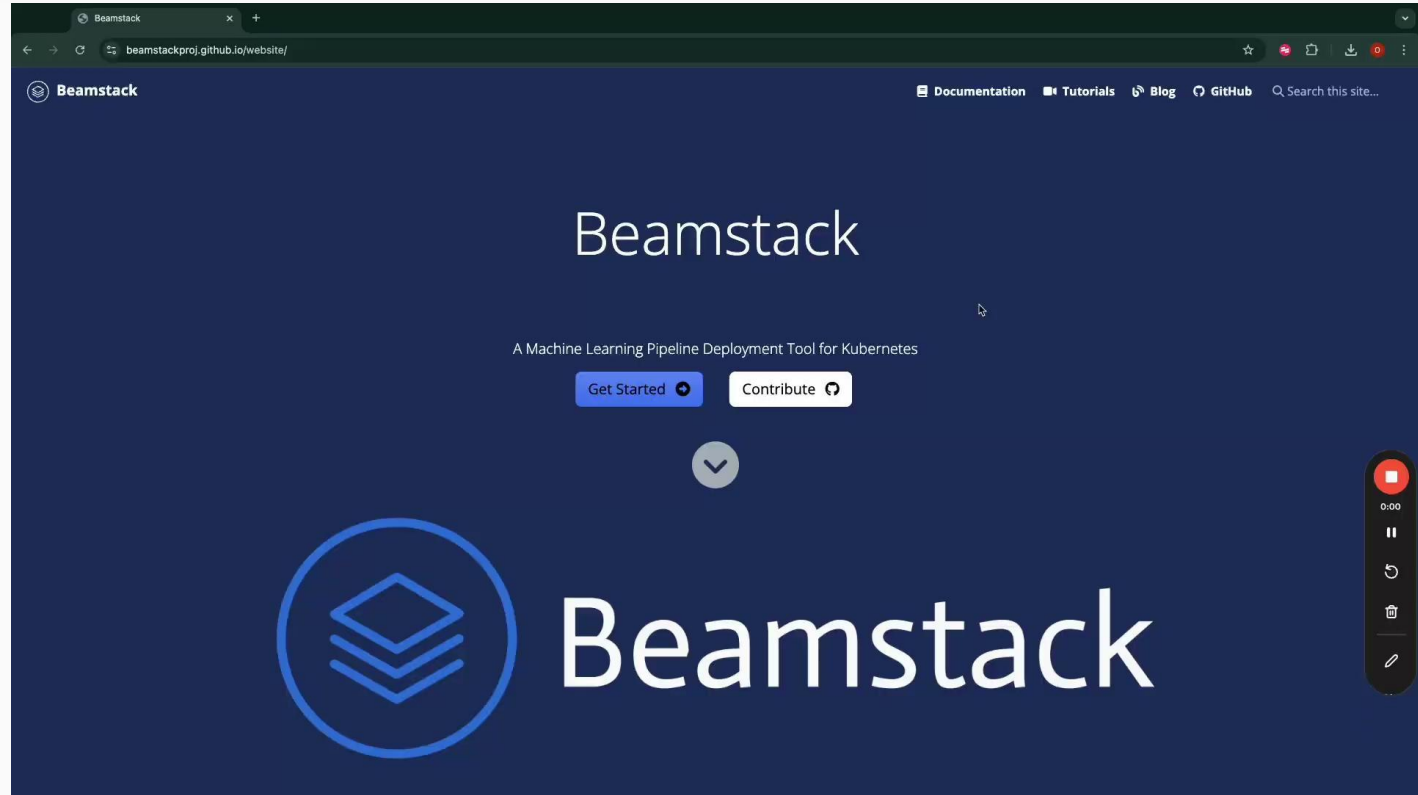
Help implement these features, as they will be detailed and listed on the Project tracker on Github



Call for Participation



How To Get Involved



Get started by
visiting our website
and going through
all of our
documentation

beamstackproj.github.io/website

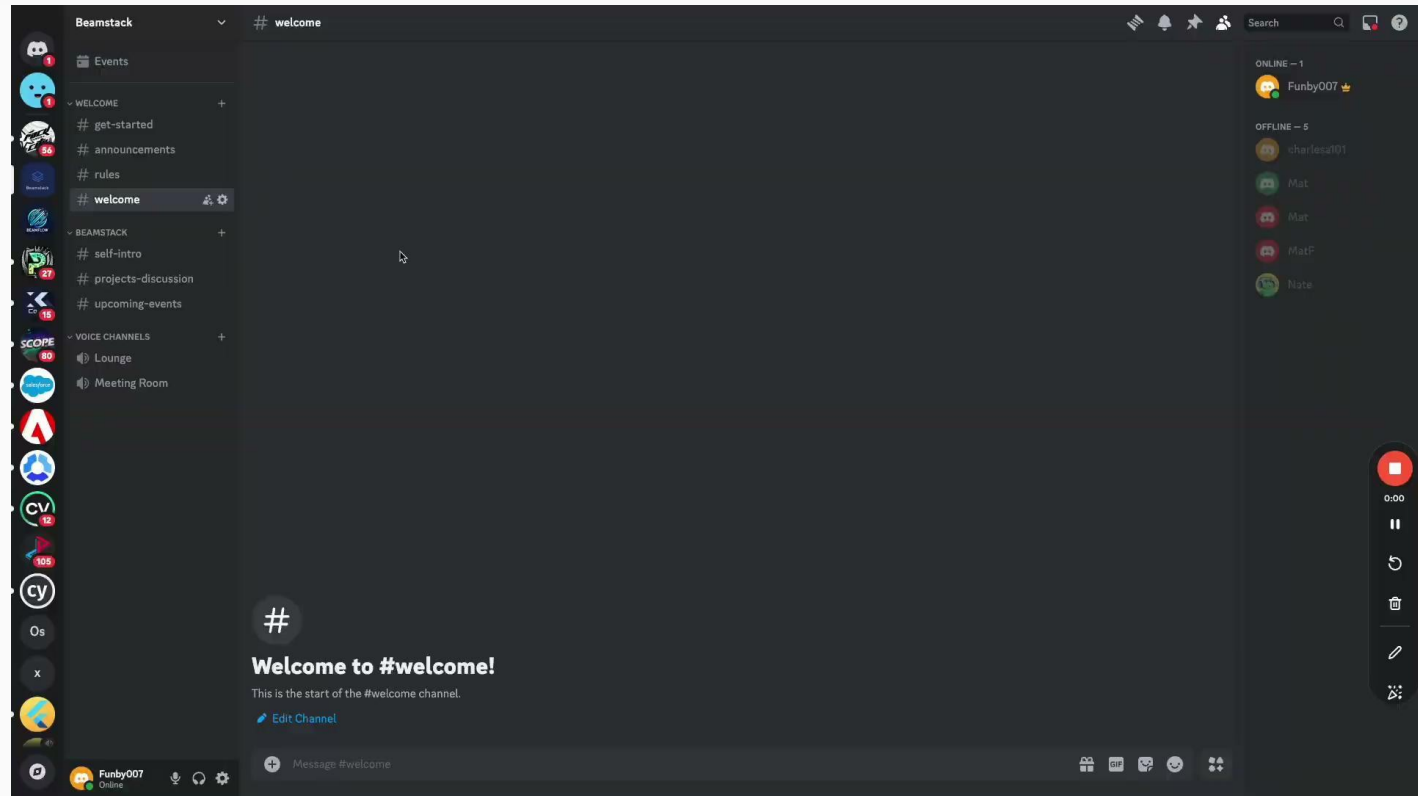


BEAM
SUMMIT

Join Our Community

Join our Discord Community today to see how you can participate and possibly contribute

QR Code on next Slide!



BEAM
SUMMIT

Thank you!



<https://github.com/beamstackproj>



<https://bit.ly/beamstack>



<https://beamstackproj.github.io/>



BEAM
SUMMIT